1-1-2015

The Neural and Behavioral Correlates of Auditory Streaming

Kate Longstaff Christison-Lagay
University of Pennsylvania, katechri@mail.med.upenn.edu

The Neural and Behavioral Correlates of Auditory Streaming

Abstract

Perceptual representations of auditory stimuli—which are called auditory streams or objects—are derived from the auditory system's ability to segregate and group stimuli based upon spectral, temporal, and spatial features. However, it remains unclear how our auditory system encodes these auditory streams at the level of the single neuron. In order to address this question directly, we first validated an animal model of auditory streaming. Specifically, we trained rhesus macaques to report their streaming percept using methodologies and controls similar to those presented in previous human studies. We found that the monkeys' behavioral reports were qualitatively consistent with those of human listeners. Next, we recorded from neurons in the primary auditory cortex while monkeys simultaneously reported their streaming percepts. We found that A1 neurons had frequency-tuned responses that habituated, independent of frequency content, as the auditory sequence unfolded over time; and we report for the first time that firing rate of A1 neurons was modulated by the monkeys' choices. This modulation increased with listening time and was independent of the frequency difference between consecutive tone bursts. Overall, our results suggest that A1 activity contributes to the sensory evidence underlying the segregation and grouping of acoustic stimuli into distinct auditory streams. However, because we observe choice-related activity based upon firing rate alone, our data are at partially at odds with Micheyl et al.'s (2005) prominent hypothesis, which argued that frequency-dependent habituation may be a coding mechanism for the streaming percept.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Neuroscience

First Advisor
Yale E. Cohen

Keywords
A1, auditory processing, macaques, perception, streaming

Subject Categories
Neuroscience and Neurobiology

THE NEURAL AND BEHAVIORAL CORRELATES OF AUDITORY STREAMING

Kate L. Christison-Lagay

A DISSERTATION

in

Neuroscience

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2015

Supervisor of Dissertation

_____

Yale Cohen, Ph.D.
Associate Professor of Otorhinolaryngology

Graduate Group Chairperson

_____

Joshua Gold, Ph.D.
Professor of Neuroscience

Dissertation Committee:

David Brainard, Ph.D., RRL Professor of Psychology
Joshua Gold, Ph.D., Professor of Neuroscience
Alan Stocker, Ph.D., Assistant Professor of Psychology
Mitchell Steinschneider, M.D., Ph.D., Professor of Neurology and Neuroscience

THE NEURAL AND BEHAVIORAL CORRELATES OF AUDITORY STREAMING

COPYRIGHT

2015

Kate Longstaff Christison-Lagay

To Sean.

To my family.

And to Gus, Sam and Hobbes.

ACKNOWLEDGMENTS

ABSTRACT

THE NEURAL AND BEHAVIORAL CORRELATES OF AUDITORY STREAMING

Kate L. Christison-Lagay

Yale E. Cohen

Perceptual representations of auditory stimuli—which are called auditory streams or objects—are derived from the auditory system's ability to segregate and group stimuli based upon spectral, temporal, and spatial features. However, it remains unclear how our auditory system encodes these auditory streams at the level of the single neuron. In order to address this question directly, we first validated an animal model of auditory streaming. Specifically, we trained rhesus macaques to report their streaming percept using methodologies and controls similar to those presented in previous human studies. We found that the monkeys' behavioral reports were qualitatively consistent with those of human listeners. Next, we recorded from neurons in the primary auditory cortex while monkeys simultaneously reported their streaming percepts. We found that A1 neurons had frequency-tuned responses that habituated, independent of frequency content, as the auditory sequence unfolded over time; and we report for the first time that firing rate of A1 neurons was modulated by the monkeys' choices. This modulation increased with listening time and was independent of the frequency difference between consecutive tone bursts. Overall, our results suggest that A1 activity contributes to the sensory evidence underlying the segregation and grouping of acoustic stimuli into distinct auditory streams. However, because we observe choice-related activity based upon firing rate alone, our data are at partially at odds with Micheyl et al.'s (2005) prominent hypothesis, which argued that frequency-dependent habituation may be a coding mechanism for the streaming percept.

TABLE OF CONTENTS

LIST OF TABLES

1   LIST OF ILLUSTRATIONS

1   CHAPTER 1: INTRODUCTION

1.1   OVERVIEW

Imagine, for a moment, that you are at a cocktail party. You are surrounded in a sea of sound: music plays in the background; your conversation partner is telling you a story; the person behind you is noisily eating tortilla chips; somewhere in the room, a group of fellow party-goers are engaged in a lively debate; and a cell phone is ringing. Each of these sound sources (e.g., the phone, the stereo speaker, the voices) produces an acoustic stimulus that happens in close temporal and spatial proximity to one another and likely has many similar frequency components. These acoustic stimuli reach your ears as an unlabeled mixture, and somehow, you are readily— and for normal listeners, seemingly effortlessly—able to segregate this mixture into distinct sounds. But how is our auditory system able to transform this enmeshed mixture of acoustic information into these distinct perceptual representations (i.e., sounds, such as the music or the cell phone's ring)?

A fundamental component of this transformation is the auditory system's ability to detect, extract, segregate, and group the spatial, spectral, and temporal regularities in the acoustic environment into distinct perceptual units (Bizley et al., 2009a; Bizley et al., 2013a; Bizley et al., 2013b; Bregman, 1990; McDermott, 2009; Russ et al., 2008b; Shinn-Cunningham, 2008; Sussman et al., 2005; Tsunada et al., 2011a; Tsunada et al., 2012; Winkler et al., 2009). In auditory neuroscience, discrete perceptual units are often called *auditory objects*, and multiple auditory objects that are grouped over time are called *auditory streams*; in common parlance, both can be called sounds (Bregman, 1990; McDermott, 2009; Shinn-Cunningham, 2008; Sussman et al., 2005; Winkler et al., 2009). These auditory perceptual units are not necessarily discrete: they can span multiple acoustic events that unfold over time (Bizley et al., 2013a;

1

Bregman, 1990; Fishman et al., 2004; Fishman et al., 2001a; Micheyl et al., 2005; Micheyl et al., 2007; Sussman et al., 2007). This enables listeners to follow the individual musical notes that form a song or to hear the sound of someone walking as 'footsteps'.

The study of auditory perception (sometimes also called auditory scene analysis (Bregman, 1990)) can be broadly categorized into two complementary approaches: a psychophysical approach that tackles the acoustic and temporal principles underlying a listener's ability to group and segregate auditory stimuli into discrete sounds; and a second that examines how the brain instantiates the above-stated principles. The latter approach can be further divided based on the scale of neural processing and has been studied from the level of the single-cell recordings up through whole-brain imaging. Below, we discuss both approaches as they relate to both the behavioral and neural correlates of auditory streaming.

1.2   REGULARITIES AND STREAM FORMATION: PSYCHOPHYSICS

Before addressing how the brain encodes auditory objects and streams, it is helpful to understand the types of acoustic and temporal cues that lead to the formation of auditory perceptions. Bregman's (1990) theory suggests that auditory percepts are formed by detecting and grouping the spectrotemporal *regularities* (e.g., harmonicity, spatial location, etc.) in the acoustic environment. That is, the brain assumes that acoustic features that are harmonically related, occur at the same location, have close temporal proximity etc. are likely to have arisen from the same sound source and should be grouped together and represented as a single distinct 'sound'. (Bregman, 1990; Grimault et al., 2002; Hill et al., 2011; Singh, 1987; van Noorden, 1975; Vliegen et al., 1999)). In contrast, dissimilar features (e.g., ones not harmonically related or from different locations) should be segregated and heard as two or more distinct sounds.

An excellent and simple example of how listeners use regularities to group (and

segregate) acoustic information is the acoustical stimulus that is discussed in Chapters 2 and 3 of this thesis, sometimes known as the galloping tones paradigm or the ABA paradigm (Bregman, 1990; Carlyon et al., 2001; Cusack, 2005; Elhilali et al., 2009; Micheyl et al., 2005; Micheyl et al., 2007). Typically, this stimulus is composed of an interleaved (asynchronous) sequence of tone bursts at two frequencies ('tone A' and 'tone B'; Fig. 3-1-1). Listeners hear this stimulus in one of two ways: (1) with all of the tone bursts grouped into one stream that sounds like a galloping rhythm; or (2) with the different frequency tone bursts segregated, and, thus, eliciting the percept of two distinct auditory streams.

Interestingly, the likelihood of hearing one or two auditory streams can be titrated by systematically varying the acoustic properties of this sequence. For example, when the frequency difference between the tone bursts in this auditory sequence is small (e.g., ≤1 semitone difference), listeners reliably report hearing one stream. On the other hand, when the frequency difference between these tone-burst sequences is large (e.g., ≥10 semitones), listeners reliably report hearing two separate streams. When the frequency difference is intermediate between these two extremes, listeners reports vary on a trial-by-trial basis (and, indeed, within trials as well (Micheyl et al., 2007)). The amount of time that a listener hears a sequence will also influence his/her reports. When listening for a short time, listeners are more likely to report a sequence with an intermediate frequency difference as one stream, but with further listening, they are more likely to report two streams (Bregman, 1990; Cusack et al., 2004; Micheyl et al., 2005). Finally, the temporal proximity of the tone bursts (i.e., whether tones are played synchronously or asynchronously) also affects a listener's choices. When the tone bursts are presented synchronously (Fig. 3-1-1, insert), instead of asynchronously as described above, listeners report hearing one stream, regardless of the frequency differences between the two tones (Elhilali et al.,

3

2009).

Although this auditory sequence has been used extensively in human studies to study the psychophysical mechanisms of human audition, there have not been any studies that explicitly tested streaming abilities of non-human animals using techniques comparable to those used in humans. Therefore, it remains unclear whether non-human animals, in fact, perceive streams in the same way as humans. We address this question in Chapter 2 by training rhesus monkeys to report their streaming percepts using a task and conditions comparable to those used in human psychophysical studies. We found that monkeys' behavioral reports were qualitatively consistent with those of human listeners.

Because the streaming task using monkeys was validated as a behavioral model, we were able to use the task to study the neural coding that underlies this behavior, and more generally, auditory perception. The next sections provide an introduction to what is known about the auditory processing of perceptual information, and previous work studying the neural correlates of auditory streaming.

1.3 A NEURAL PATHWAY FOR AUDITORY PERCEPTION: THE VENTRAL AUDITORY PATHWAY

How and where does perception occur in the auditory system? Correlates of auditory perception can be found as early as the cochlear nucleus (Pressnitzer et al., 2001; Pressnitzer et al., 2008), and stimulus-specific adaptation (a proposed mechanism for encoding auditory streams) is found in the auditory thalamus (Anderson et al., 2009; Antunes et al., 2010) and auditory cortex (Szymanski et al., 2009; Taaseh et al., 2011; Xu et al., 2014). However, because we are interested in studying the neural correlates of perception, the following discussion will focus on the contribution of the cortex to streaming and, in particular, the contribution of the 'ventral' auditory pathway.

The ventral auditory pathway is one of two pathways that are generally thought to process auditory information in the cortex; and it is thought to process a sound's identity, content and meaning (consequently, it is sometimes referred to as the 'what' pathway) (Kaas et al., 1999; Romanski et al., 1999). The other pathway, the 'dorsal' pathway, contributes to sound localization and audiomotor action. We should note that this parsing of the auditory brain is not universally accepted and other variants have been proposed (Griffiths, 2008; Rauschecker, 2012; Rauschecker et al., 2009). Because the question of the grouping and segregation of acoustic information into streams is a question of sound *identity*, the ventral pathway is the more obvious pathway to target for our initial study; it remains an open question, though, whether and how the dorsal pathway might contribute to auditory perception in those situations when stimuli can be segregated using spatial information.

In the rhesus macaque, the ventral pathway begins in the core auditory fields, primary auditory cortex (A1) and field R; Chapter 3 describes recordings from A1 in monkeys that are reporting streaming percepts. The core areas project to the anterolateral (AL) and middle-lateral belt regions of the auditory cortex (Kaas et al., 2000; Rauschecker et al., 2000), which, in turn, project directly and indirectly to the ventrolateral prefrontal cortex (vlPFC) (Romanski et al., 1999). Although only A1 was targeted in the current study, these other regions represent appealing future recording targets to study the transformation of neural activity during the streaming task along the entire ventral auditory pathway.

There is no universal consensus on what information is coded in each region of the ventral auditory pathway. In fact, there remains a great deal of debate over even what acoustic features are preferentially processed in each of these regions, let alone the contribution of these regions to 'higher order' processing, such as categorization or choice. Nonetheless, it is thought

that neurons in this pathway encode increasingly more complex attributes of a stimulus the further along the pathway one progresses. Generally speaking, neurons in the core auditory fields seem to be sensitive to a number of low-level acoustic features, such as frequency, intensity, and location, as well as some more derived properties, such as timbre and stimulus novelty (Bendor et al., 2005; Bizley et al., 2009a; Bizley et al., 2010; Bizley et al., 2009b; Bizley et al., 2013b; Javitt et al., 1994; Razak, 2011; Schebesch et al., 2010 ; Ulanovsky et al., 2004; Versnel et al., 1998 ; Wang et al., 1995; Watkins et al., 2011; Werner-Reiss et al., 2008; Zhou et al., 2010). Further along, AL neurons respond preferentially to band-pass noise, frequency-modulated sweeps, and vocalizations (Christison-Lagay et al., 2014b; Kikuchi et al., 2010; Rauschecker et al., 2000; Rauschecker et al., 2004; Rauschecker et al., 1995; Tian et al., 2004; Tian et al., 2001; Tsunada et al., 2011a). The auditory belt and parabelt regions show an even greater degree of stimulus selectivity, such as selectivity for vocalizations (Chang et al., 2010; Leaver et al., 2010; Obleser et al., 2009; Obleser et al., 2010; Obleser et al., 2006). Additionally, other nearby auditory fields show a preference for voices (Perrodin et al., 2011; Petkov et al., 2008). Finally, vlPFC neurons are modulated more by the cognitive components of audition, such as non-spatial auditory attention, auditory working memory, and the referential meaning of vocalizations (Cohen et al., 2009c; Gifford III et al., 2005b; Lee et al., 2009; Ng et al., 2013; Plakke et al., 2013; Plakke et al., 2015; Russ et al., 2008a; Russ et al., 2008b).

As previously mentioned, there is still a great degree of controversy about where choice-modulated neural activity emerges in the pathway. For example, some studies of A1 neurons have found that neural activity correlates with a monkey's reports of category identify (Selezneva et al., 2006), pitch (Bizley et al., 2013b) and amplitude modulation (Niwa et al., 2012b). Other studies have suggested that A1 may contain choice-related activity pertinent to streaming (see the

next section, Regularities and Stream formation: neural basis for discussion; (Elhilali et al., 2009; Micheyl et al., 2005)).

However, there is another body of literature that suggests that neural correlates of perception are not found until later portions of the ventral pathway (either later areas of the auditory cortex (Chang et al., 2010; Gutschalk et al., 2008; Mesgarani et al., 2012), or the vlPFC (Lee et al., 2009; Russ et al., 2008a; Tsunada et al., 2011b)). MEG data, for example, suggest that the neural correlates of a listener hearing a sound, while engaged in an informational-masking paradigm, are found in the secondary (belt) auditory cortex (Gutschalk et al., 2008); and correlates of perceptual judgments about communication sounds (species-specific vocalizations and speech sounds) have also been found in belt region of the auditory cortex and higher auditory cortices (Chang et al., 2010; Mesgarani et al., 2012). In recent studies of phonemic categorization, although neural correlates of categorization *were* found in higher auditory cortex, perceptual judgments were not. Instead, choice-related activity emerged at the level of the vlPFC (Lee et al., 2009; Russ et al., 2008a; Tsunada et al., 2011a).

It is unclear why some studies find choice-related activity as early as the core auditory cortex and others do not find it until much further downstream. However, it is possible that the choice and complexity of stimuli and task contribute to the difference. The studies that find choice-related activity in A1 used relatively simple tasks and/or stimuli (such as the discriminating pitch or depth of amplitude modulation); these stimuli can be represented directly in the firing rates of A1 neurons. However, choice activity attendant to tasks that use complex stimuli, such as vocalizations, or tasks in which the decisions is based on more derived stimulus properties is not seen in the auditory cortex(Tsunada et al., under review). Thus, choice-related activity may emerge where neurons are able to represent sensory evidence relevant to the choice;

and therefore, choice-related activity may originate in more than just one area. This conclusion assumes that these choice signals represent a feed-forward process and do not reflect feedback from higher-decision areas (Nienborg et al., 2014).

1.4 REGULARITIES AND STREAM FORMATION: NEURAL BASIS

As discussed earlier, there has been extensive psychophysical work using the streaming task with humans. However, thus far, the study of the neural correlates of streaming have been studied only in A1 using passive-listening paradigms, or paradigms in which the monkeys were not required to report streaming percepts (Fishman et al., 2004; Fishman et al., 2001a; Micheyl et al., 2005). These studies have shown that in response to alternating tone sequences, neurons in A1 responses adapt to tones over time as a function of its frequency and repetition rate. Specifically, A1 neurons have been reported to respond more to their best frequency (alternately defined as the frequency that elicits the highest response at a fixed intensity, or the frequency that elicits a reliable response at the lowest intensity) and are less suppressed by repeated presentations of this frequency than this 'non-best' frequency (defined as frequencies away from the best frequency) (Fishman et al., 2004; Fishman et al., 2001a; Micheyl et al., 2005).

This pattern of A1 activity is consistent with the hypothesis that stream segregation is represented in a place code (Eggermont, 2001; Steinschneider et al., 1990); see Fig 1-2. In this theory, place along A1's tonotopic map would encode the number of perceived streams perceived as a function of the spatial separation between active neural populations: one stream is perceived when there is one peak of activity, whereas two streams are perceived when there are two discernable peaks of activity. A place code could effectively use differential rates of habituation to encode the number of streams: neurons habituated more to non-best frequencies, and therefore, after repeated presentations of a tone, neurons may respond robustly *only* to their best

frequencies. This leads to two distinct neural populations on A1's tonotopic map, with each population responding only to its best frequency, and each encoding a separate stream. It is important to note that the term 'place code' here refers to the neural population's combination of spatial and rate codes, and does not imply that there is a 'labeled line' or specific place that encodes 1 versus 2 streams in the primary auditory cortex.

This kind of neural place code, however, is insufficient to explain other aspects of streaming. As described earlier, Elhilali et al. (2009) found that when the tone bursts are presented synchronously—instead of asynchronously as is typical with studies of auditory streaming—listeners report hearing one stream. This is observation seems at odds with this neural-place coding hypothesis. Because of this discrepancy, Elhilali et al. (2009) proposed a different model of stream segregation in which the timing of activity encodes streams: their hypothesis argues that streams are formed on the basis of the detection of neural populations with temporally coherent activity. Thus, for both synchronous tone sequences or alternating sequences with small frequency separations, the active neural population(s) would respond simultaneously, which would be read out downstream as evidence for a single stream. On the other hand, tone sequences with large frequency separations produce two neural populations responding at different times, and would be interpreted as two distinct auditory streams.

Likely, both the neuronal arrangement (e.g., topographical/tonotopic) and temporal pattern of activity play roles in stream formation. However, a strict interpretation of temporal coherence is also likely insufficient, as recent studies have found that temporally coherent sounds can, in fact, be segregated into multiple, discrete streams under certain conditions (Micheyl et al., 2010; Micheyl et al., 2013a; Micheyl et al., 2013b). Although the current work does not directly address either of these models, the results of the study presented in Chapter 3 support a mechanism to

9

encode streams that would use a combination of place and temporal dynamics.

1.5 CONCLUSIONS

In spite of the progress that has been made in the study of the neural mechanisms involved in auditory streaming, the neural code underlying the relationship between the acoustic features of an auditory stimulus, neural activity, and the listener's percept remains unclear. Ultimately, neither psychophysical studies nor human-imaging studies can provide sufficient insight into the neural code underlying perceptual processing: the relationship between perception and neural activity can only be evaluated by directly testing both simultaneously. Although previous studies have provided a great deal of insight into how the brain encodes acoustic stimuli, few studies have directly and systematically tested neural activity using the same behavioral tests and stimuli used in humans. The studies described in Chapters 2 and 3 aimed to address this gap in the literature by (1) directly testing whether non-human animals stream sounds in a manner consistent with humans and (2) studying the way in which A1 neurons encode both the acoustic and behavioral aspects of the task.
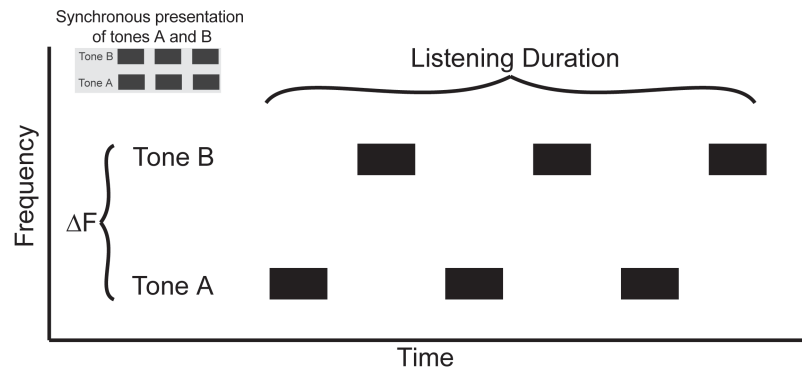
Figure 1-1: Schematic of the auditory stimulus to test auditory streaming.
The auditory stimulus is an asynchronous sequence of two types of tone bursts: tone A and tone B. Typically, tones A and B were presented asynchronously but were at times presented simultaneously (see inset at upper left). Small frequency differences (<1 semitone), short listening durations, and synchronous tone presentation bias listeners towards perceiving one stream; larger frequency (>10 semitones) and long listening durations bias listeners towards perceiving two streams. The units on the x- and y-axes are arbitrary.
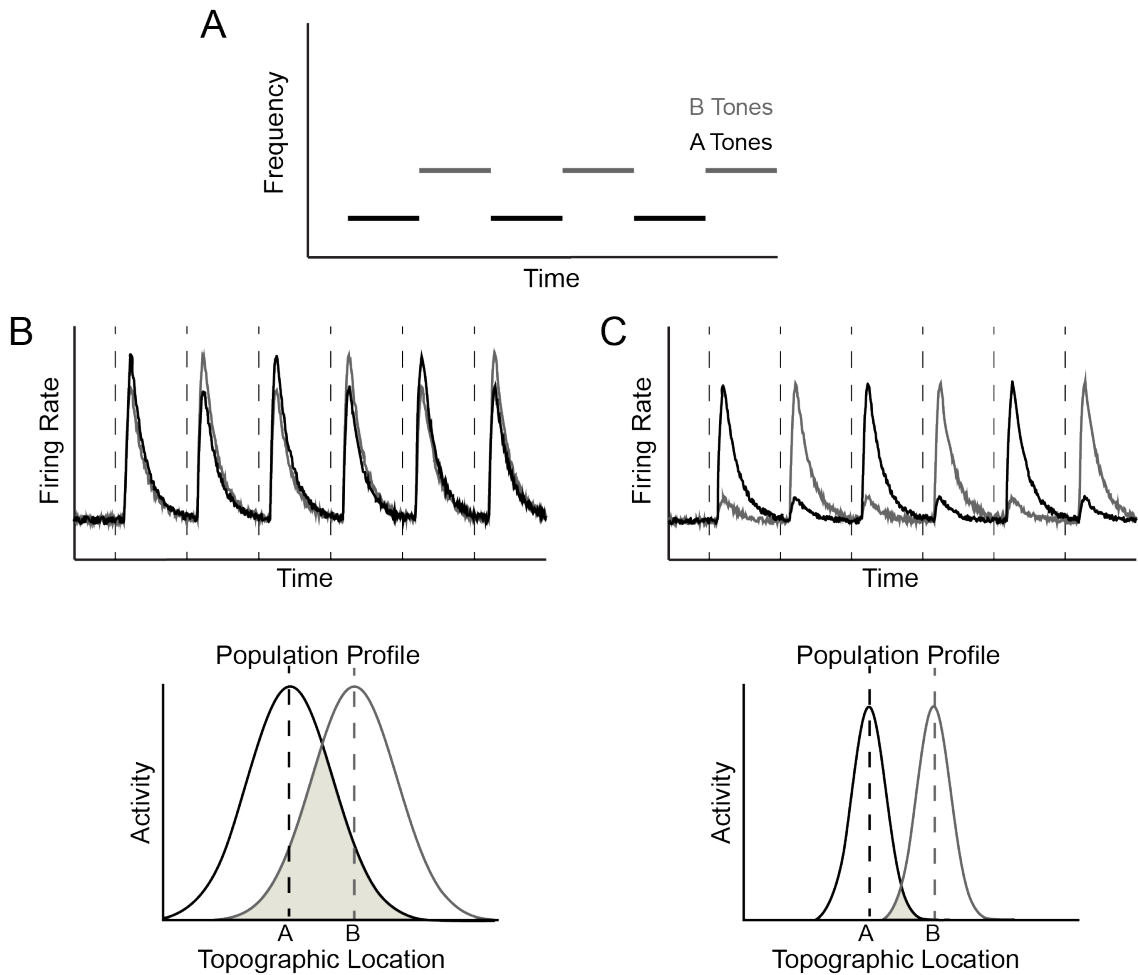
Figure 1-2: Putative neural mechanism mediating auditory streaming.
Panel A presents an example of an alternating tone sequence as used in the streaming task. Tones at two frequencies (A and B) are presented in an alternating fashion; here, the frequency different should be considered 'intermediate'. **Panel B** shows an example of the neural response to a streaming sequence with intermediate frequency separation early in the sequence presentation. The top row shows the response of two neurons, with best frequencies at either tone A's frequency (shown in black) or tone B's frequency (shown in gray). Early in the sequence presentation, both neurons respond robustly to both frequencies. This is shown schematically in the bottom row: the filled gray area represents the topographic locations in A1 that would respond to both tone A and tone B frequencies. **Panel C** shows the neural response later in the same trial. The top row shows the response of same two neurons after frequency-specific habituation has occurred. There is still a robust response to the neuron's best frequency, but each neuron has stopped responding to the other frequency. The bottom row shows schematically how frequency-specific habituation reduces the area of cortex responding to a given frequency, and leads to two separate populations of neurons that encode the stimulus. Modified from Christison-Lagay et al. (2015).

1.7 REFERENCES

Anderson, L.A., Christianson, G.B., Linden, J.F. 2009. Stimulus-specific adaptation occurs in the auditory thalamus. J Neurosci 29, 7359-63.

Antunes, F.M., Nelken, I., Covey, E., Malmierca, M.S. 2010. Stimulus-specific adaptation in the auditory thalamus of the anesthetized rat. PLoS One 5, e14071.

Bendor, D., Wang, X. 2005. The neuronal representation of pitch in primate auditory cortex. Nature 436, 1161-5.

Bizley, J.K., Walker, K.M. 2009a. Distributed sensitivity to conspecific vocalizations and implications for the auditory dual stream hypothesis. J Neurosci 29, 3011-3013.

Bizley, J.K., Cohen, Y.E. 2013a. The what, where and how of auditory-object perception. Nat Rev Neurosci 14, 693-707.

Bizley, J.K., Walker, K.M., King, A.J., Schnupp, J.W. 2010. Neural ensemble codes for stimulus periodicity in auditory cortex. J Neurosci 30, 5078-91.

Bizley, J.K., Walker, K.M., Silverman, B.W., King, A.J., Schnupp, J.W. 2009b. Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. J Neurosci 29, 2064-75.

Bizley, J.K., Walker, K.M., Nodal, F.R., King, A.J., Schnupp, J.W. 2013b. Auditory cortex represents both pitch judgments and the corresponding acoustic cues. Curr Biol 23, 620-5.

Bregman, A.S. 1990. Auditory Scene Analysis: The Perceptual Organization of Sound MIT Press, Cambridge, MA.

Carlyon, R.P., Cusack, R., Foxton, J.M., Robertson, I.H. 2001. Effects of Attention and Unilateral Neglect on Auditory Stream Segregation. J Exp Psychol Hum Percept Perform 27, 115-127.

Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., Knight, R.T. 2010. Categorical speech representation in human superior temporal gyrus. Nat Neurosci 13, 1428-32.

Christison-Lagay, K.L., Gifford, A.M., Cohen, Y.E. 2015. Neural correlates of auditory scene analysis and perception. International journal of psychophysiology : official journal of the International Organization of Psychophysiology 95, 238-45.

Christison-Lagay, K.L., Bennur, S., Blackwell, J., Lee, J.H., Schroeder, T., Cohen, Y.E. 2014b. Natural variability in species-specific vocalizations constrains behavior and neural activity. Hear Res 312, 128-42.

Cohen, Y.E., Russ, B.E., Davis, S.J., Baker, A.E., Ackelson, A.L., Nitecki, R. 2009c. A functional role for the ventrolateral prefrontal cortex in non-spatial auditory cognition. Proc Natl Acad Sci U S A 106, 20045-50.

Cusack, R. 2005. The Intraparietal Sulcus and Perceptual Organization. J Cogn Neurosci 17, 641-651.

Cusack, R., Deeks, J., Aikman, G., Carlyon, R.P. 2004. Effects of Location, Frequency Region, and Time Course of Selective Attention on Auditory Scene Analysis. Journal of Experimental Psychology: Human Perception and Performance 30, 643-656.

Eggermont, J.J. 2001. Between sound and perception: reviewing the search for a neural code. Hear Res, 1-42.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A.J., Shamma, S.A. 2009. Temporal coherence in the

perceptual organization and cortical representation of auditory scenes. Neuron 61, 317-29.

Fishman, Y.I., Arezzo, J.C., Steinschneider, M. 2004. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J Acoust Soc Am 116, 1656-70.

Fishman, Y.I., Reser, D.H., Arezzo, J.C., Steinschneider, M. 2001a. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear Res 151, 167-187.

Gifford III, G.W., MacLean, K.A., Hauser, M.D., Cohen, Y.E. 2005b. The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. Journal of Cognitive Neuroscience 17, 1471-1482.

Griffiths, T.D. 2008. Sensory Systems: Auditory Action Streams? Curr Biol 18, R387-R388.

Grimault, N., Bacon, S.P., Micheyl, C. 2002. Auditory stream segregation on the basis of amplitude-modulation rate. J Acoust Soc Am 111, 1340.

Gutschalk, A., Micheyl, C., Oxenham, A.J. 2008. Neural Correlates of Auditory Perceptual Awareness under Informational Masking. Plos Biol 6, e138.

Hill, K.T., Bishop, C.W., Yadav, D., Miller, L.M. 2011. Pattern of BOLD signal in auditory cortex relates acoustic response to perceptual streaming. BMC Neurosci 12, 85.

Javitt, D.C., Steinschneider, M., Schroeder, C.E., Vaughan, H.G., Jr., Arezzo, J.C. 1994. Detection of stimulus deviance within primate primary auditory cortex: intracortical mechanisms of mismatch negativity (MMN) generation. Brain Res 667, 192-200.

Kaas, J., Hackett, T. 2000. Subdivisions of Auditory Cortex and Processing Streams in Primates. Proceedings of the National Academy of Sciences of the United States of America 97, 11793-11799.

Kaas, J.H., Hackett, T.A. 1999. 'What' and 'where' processing in auditory cortex. Nat Neurosci 2, 1045-1047.

Kikuchi, Y., Horwitz, B., Mishkin, M. 2010. Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. J Neurosci 30, 13021-30.

Leaver, A.M., Rauschecker, J.P. 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J Neurosci 30, 7604-12.

Lee, J.H., Russ, B.E., Orr, L.E., Cohen, Y. 2009. Prefrontal activity predicts monkeys' decisions during an auditory category task. Frontiers in Integrative Neuroscience 3.

McDermott, J. 2009. The cocktail party problem. Curr Biol 19, R1024-R1027.

Mesgarani, N., Chang, E.F. 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485, 233-6.

Micheyl, C., Hunter, C., Oxenham, A.J. 2010. Auditory stream segregation and the perception of across-frequency synchrony. J Exp Psychol Hum Percept Perform 36, 1029-1039.

Micheyl, C., Tian, B., Carlyon, R.P., Rauschecker, J.P. 2005. Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron 48, 139-48.

Micheyl, C., Kreft, H., Shamma, S.A., Oxenham, A.J. 2013a. Temporal coherence versus harmonicity in auditory stream formation. J Acoust Soc Am 133, EL188.

Micheyl, C., Hanson, C., Demany, L., Shamma, S.A., Oxenham, A.J. 2013b. Auditory Stream Segregation for Alternating and Synchronous Tones. J Exp Psychol Hum Percept Perform.

Micheyl, C., Carlyon, R.P., Gutschalk, A., Melcher, J.R., Oxenham, A.J., Rauschecker, J.P., Tian, B., Courtenay Wilson, E. 2007. The role of auditory cortex in the formation of auditory
14

streams. Hear Res 229, 116-31.

Ng, C.W., Plakke, B., Poremba, A. 2013. Neural correlates of auditory recognition memory in the primate dorsal temporal pole. J Neurophysiol.

Nienborg, H., Cumming, B.G. 2014. Decision-related activity in sensory neurons may depend on the columnar architecture of cerebral cortex. J Neurosci 34, 3579-85.

Niwa, M., Johnson, J.S., O'Connor, K.N., Sutter, M.L. 2012b. Activity related to perceptual judgment and action in primary auditory cortex. J Neurosci 32, 3193-210.

Obleser, J., Eisner, F. 2009. Pre-lexical abstraction of speech in the auditory cortex. Trends Cogn. Sci. 13, 14-9.

Obleser, J., Leaver, A.M., Vanmeter, J., Rauschecker, J.P. 2010. Segregation of vowels and consonants in human auditory cortex: evidence for distributed hierarchical organization. Frontiers in psychology 1, 232.

Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roettinger, M., Eulitz, C., Rauschecker, J.P. 2006. Vowel sound extraction in anterior superior temporal cortex. Hum Brain Mapp 27, 562-71.

Perrodin, C., Kayser, C., Logothetis, N.K., Petkov, C.I. 2011. Voice cells in the primate temporal lobe. Curr Biol 21, 1408-15.

Petkov, C.I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., Logothetis, N.K. 2008. A voice region in the monkey brain. Nat Neurosci 11, 367-74.

Plakke, B., Diltz, M.D., Romanski, L.M. 2013. Coding of vocalizations by single neurons in ventrolateral prefrontal cortex. Hear Res 305, 135-43.

Plakke, B., Hwang, J., Romanski, L.M. 2015. Inactivation of Primate Prefrontal Cortex Impairs Auditory and Audiovisual Working Memory. J Neurosci 35, 9666-75.

Pressnitzer, D., Meddis, R., Delahaye, R., Winter, I.M. 2001. Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus. J Neurosci 15, 6377-6386.

Pressnitzer, D., Sayles, M., Micheyl, C., Winter, I.M. 2008. Perceptual organization of sound begins in the auditory periphery. Curr Biol 18, 1124-8.

Rauschecker, J.P. 2012. Ventral and dorsal streams in the evolution of speech and language. Frontiers in evolutionary neuroscience 4, 7.

Rauschecker, J.P., Tian, B. 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci U S A 97, 11800-6.

Rauschecker, J.P., Tian, B. 2004. Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. J Neurophysiol 91, 2578-89.

Rauschecker, J.P., Scott, S.K. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12, 718-24.

Rauschecker, J.P., Tian, B., Hauser, M. 1995. Processing of complex sounds in the macaque nonprimary auditory cortex. Science 268, 111-4.

Razak, K.A. 2011. Systematic representation of sound locations in the primary auditory cortex. J Neurosci 31, 13848-13859.

Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., Rauschecker, J.P. 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci 2, 1131-6.

Russ, B.E., Orr, L.E., Cohen, Y.E. 2008a. Prefrontal neurons predict choices during an auditory same-different task. Curr Biol 18, 1483-8.

Russ, B.E., Ackelson, A.L., Baker, A.E., Cohen, Y.E. 2008b. Coding of auditory-stimulus identity in the auditory non-spatial processing stream. J Neurophysiol 99, 87-95.

Schebesch, G., Lingner, A., Firzlaff, U., Wiegrebe, L., Grothe, B. 2010. Perception and neural representation of size-variant human vowels in the Mongolian gerbil (Meriones unguiculatus). Hearing research 261, 1-8.

Selezneva, E., Scheich, H., Brosch, M. 2006. Dual Time Scales for Categorical Decision Making in Auditory Cortex. Curr Biol 16, 2428-2433.

Shinn-Cunningham, B.G. 2008. Object-based auditory and visual attention. Trends Cogn Sci 12, 182-6.

Singh, P.G. 1987. Perceptual organization of compex-tone sequences: A tradeoff between pith and timbre? J Acoust Soc Am 82, 886-889.

Steinschneider, M., Arezzo, J.C., Vaughan, J., Herbert G. 1990. Tonotopic features of speech-evoked activity in primate auditory cortex. Brain Res 519, 158-168.

Sussman, E.S., Bregman, A.S., Wang, W.J., Khan, F.J. 2005. Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. Cognitive Affective & Behavioral Neuroscience 5, 93-110.

Sussman, E.S., Horvath, J., Winkler, I., Orr, M. 2007. The role of attention in the formation of auditory streams. Perception & Psychophysics 69, 136-152.

Szymanski, F.D., Garcia-Lazaro, J.A., Schnupp, J.W. 2009. Current source density profiles of stimulus-specific adaptation in rat auditory cortex. J Neurophysiol 102, 1483-90.

Taaseh, N., Yaron, A., Nelken, I. 2011. Stimulus-specific adaptation and deviance detection in the rat auditory cortex. PLoS One 6, e23369.

Tian, B., Rauschecker, J.P. 2004. Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. J Neurophysiol 92, 2993-3013.

Tian, B., Reser, D., Durham, A., Kustov, A., Rauschecker, J.P. 2001. Functional specialization in rhesus monkey auditory cortex. Science 292, 290-3.

Tsunada, J., Lee, J.H., Cohen, Y.E. 2011a. Representation of speech categories in the primate auditory cortex. J Neurophysiol.

Tsunada, J., Lee, J.H., Cohen, Y.E. 2012. Differential representation of auditory categories between cell classes in primate auditory cortex. J Physiol 590, 3129-39.

Tsunada, J., Liu, A.S., Gold, J.I., Cohen, Y.E. under review. Causal contribution of primate auditory cortex to auditory perceptual decision-making.

Tsunada, J., Baker, A.E., Christison-Lagay, K.L., Davis, S.J., Cohen, Y.E. 2011b. Modulation of cross-frequency coupling by novel and repeated stimuli in the primate ventrolateral prefrontal cortex. Front Psychology 2.

Ulanovsky, N., Las, L., Farkas, D., Nelken, I. 2004. Multiple time scales of adaptation in auditory cortex neurons. J Neurosci 24, 10440-10453.

van Noorden, L. 1975. Temporal coherence in the perception of tone sequences. Technical University Eindhoven. pp. 1-127.

Versnel, H., Shamma, S.A. 1998. Spectral-ripple representation of steady-state vowels in primary auditory cortex. The Journal of the Acoustical Society of America 103, 2502-14.

Vliegen, J., Oxenham, A.J. 1999. Sequential stream segregation in the absence of spectral cues. J Acoust Soc Am 105, 339-346.

Wang, X.Q., Merzenich, M.M., Beitel, R., Schreiner, C.E. 1995. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: Temporal and spectral characteristics. Journal of Neurophysiology 74, 2685-2706.

Watkins, P.V., Barbour, D.L. 2011. Level-tuned neurons in primary auditory cortex adapt differently to loud versus soft sounds. Cerebral Cortex 21, 178-190.

Werner-Reiss, U., Groh, J.M. 2008. A rate code for sound azimuth in monkey auditory cortex:

implications for human neuroimaging studies. J Neurosci 28, 3747-3758.

Winkler, I., Denham, S.L., Nelken, I. 2009. Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci 13, 532-40.

Xu, X., Yu, X., He, J., Nelken, I. 2014. Across-ear stimulus-specific adaptation in the auditory cortex. Frontiers in neural circuits 8, 89.

Zhou, Y., Wang, X.Q. 2010. Cortical processing of dynamic sound envelope transitions. J Neurosci 30, 16741-16754.

2 CHAPTER 2: BEHAVIORAL CORRELATES OF AUDITORY STREAMING IN RHESUS MACAQUES.

2.1 ABSTRACT

Perceptual representations of auditory stimuli (i.e., sounds) are derived from the auditory system's ability to segregate and group the spectral, temporal, and spatial features of auditory stimuli—a process called 'auditory scene analysis'. Psychophysical studies have identified several of the principles and mechanisms that underlie a listener's ability to segregate and group acoustic stimuli. One important psychophysical task that has illuminated many of these principles and mechanisms is the 'streaming' task. Despite the wide use of this task to study psychophysical mechanisms of human audition, no studies have explicitly tested the streaming abilities of non-human animals using the standard methodologies employed in human-audition studies. Here, we trained rhesus macaques to participate in the streaming task using methodologies and controls similar to those presented in previous human studies. Overall, we found that the monkeys' behavioral reports were qualitatively consistent with those of human listeners, thus suggesting that this task may be a valuable tool for future neurophysiological studies.

2.2 INTRODUCTION

One of the fundamental tasks of the auditory system is to transform low-level sensory representations of acoustic stimuli into perceptual representations (i.e., sounds) that can guide behavior (Bizley et al., 2013a; Griffiths et al., 2004; Shamma et al., 2010; Shinn-Cunningham, 2008). These perceptual representations form the core building blocks of our hearing experience

(Bregman, 1990; Griffiths et al., 2004; Shamma, 2008) and are derived from the auditory system's ability to segregate and group the spectral, temporal, and spatial features of auditory stimuli—a process called 'auditory scene analysis' (Bregman, 1990; McDermott, 2009; Winkler et al., 2009). Auditory scene analysis enables a listener to follow, for example, the melody that is carried by a banjo in a band or to track a friend's voice in a noisy restaurant (McDermott, 2009; Shinn-Cunningham, 2008).

Psychophysical studies have identified several of the principles and mechanisms that underlie a listener's ability to segregate and group acoustic stimuli (Horvath et al., 2001; Rahne et al., 2009; Sussman, 2005; Sussman et al., 2007). One important psychophysical task that has illuminated many of these principles and mechanisms is the 'streaming' task (Bregman, 1990; Carlyon et al., 2001; Cusack, 2005; Elhilali et al., 2009; Micheyl et al., 2007). Typically, the streaming task is a one-interval, two-alternative forced choice task in which an auditory stimulus—composed of an interleaved sequence of tone bursts (Fig. 2-1)—is presented and a listener reports whether she heard one or two streams. By varying the spectral, temporal, and other properties of this sequence, the probability that a listener reports one or two streams is systematically altered. For example, when the frequency difference between the tone bursts in the two sequences is small (e.g., $\leq$1 semitone difference), listeners systematically report hearing one stream. On the other hand, when the frequency difference between these tone-burst sequences is large (e.g., $\geq$10 semitones), listeners systematically report hearing two separate streams. When the frequency difference is intermediate between these two extremes, the reports become less reliable: on alternating trials, listeners report hearing one or two streams.

Despite the wide use of this task (and variants of it) to study psychophysical mechanisms of human audition (Shamma et al., 2011), no studies have explicitly tested the streaming abilities

19

of non-human animals using the standard methodologies employed in human-audition studies. Instead, previous studies have indirectly tested streaming (Izumi, 2002; Ma et al., 2010; Noda et al., 2012). For example, in Ma et al. (2010), ferrets reported hearing a 'target' tone that was embedded in a tone-burst sequence. This experimental strategy to test streaming is reasonable because many non-human animals process auditory stimuli and hear sounds in a manner similar to that of human listeners (Izumi, 2002; Kuhl et al., 1975b; Kuhl et al., 1982; Miller et al., 2001; Petkov et al., 2003; Petkov et al., 2007; Recanzone et al., 2008). Consequently, it was assumed that, like humans (Elhilali et al., 2009), these ferret listeners could only detect the target tone when the auditory stimulus was segregated into two streams.

However, if the goal of testing the auditory perceptual abilities of non-human animals is to develop them as models of human-brain function, it is imperative to use methodologies and controls that are comparable to those used with human listeners so that valid inferences can be made regarding human audition and cognition. Here, we trained rhesus macaques to participate in a streaming task using methodologies and controls similar to those presented in previous human studies. Overall, we found that the monkeys' behavioral reports were consistent with those of human listeners, thus suggesting that this task may be a valuable tool for future neurophysiological studies.

2.3   EXPERIMENTAL PROCEDURE

2.3.1   Experimental Chamber

Psychophysical sessions were conducted in a darkened room with sound-attenuating walls. A monkey (*Macaca mulatta*; Monkey H or Monkey S) was seated in a primate chair in the center of the room. A touch-sensitive joystick was attached to the chair. The monkey moved the

joystick during the behavioral task to indicate his behavioral report.

2.3.2   Auditory Stimulus

The auditory stimulus was a sequence tone bursts (40-ms duration with a 5-ms $\cos^2$ ramp at a sound level of 65 dB SPL) that alternated between two types of tone bursts, called here 'tone A' and 'tone B'. The inter-tone-burst interval was 13 Hz. Auditory stimuli were generated using the RX6 digital-signal-processing platform (TDT Inc.) and were presented by a studio-monitor speaker (Yamaha MSP7).

2.3.3   Behavioral Task

The streaming task was a single-interval, two-alternative-forced-choice discrimination task that required the monkey to report whether he heard one or two streams (Fig. 2-2). A trial began with the presentation of the auditory sequence (Fig. 2-1). Following offset of the auditory stimulus, an LED was illuminated, and the monkey had 3000 ms to move the joystick (a) to the right to report one stream or (b) to the left to report two streams.

Training Procedure and Reward Structure

During the initial training sessions, tones A and B were presented at frequency differences that, in humans (Cusack, 2005; Micheyl et al., 2005), elicit reliable reports of one or two streams (i.e., ≤1.0 or ≥10 semitones, respectively). On these trials, the monkey received consistent feedback: he was only rewarded for reporting a 'correct' response. Specifically, when the frequency difference between tone A and tone B was ≤1.0 semitone, the monkey was rewarded when he moved the joystick to the right. When the frequency difference was ≥10 semitones, the monkey was rewarded when he moved the joystick to the left.

21

After the monkey's performance stabilized (i.e., they were performing significantly above chance during entire behavioral sessions), we presented auditory sequences that contained both the 'extreme' frequency differences (≤1.0 or ≥10 semitones) as well as frequency differences that were 'intermediate' between these two extremes (i.e., >1 and <10 semitones). Because stimuli with these intermediate frequency differences do not elicit reliable reports of one or two streams in human listeners (Bregman, 1990; Bregman et al., 2000; Cusack, 2005; Elhilali et al., 2009; Micheyl et al., 2007), there was not a 'correct' answer. Consequently, on these trials, the monkeys did not receive consistent feedback: they received rewards on 50% of randomly selected trials; the decision to reward was independent of their behavioral report.

Behavioral-testing Strategy

We manipulated four parameters of the tone-sequence: the frequency difference between tones A and B; the duration of the auditory sequence; the temporal relationship between tones A and B; and the frequency of tone A. These first three parameters manipulations tested whether the monkeys' reports were modulated in a manner consistent with human listeners' reports (Bregman, 1990; Elhilali et al., 2009; Micheyl et al., 2007). The last parameter manipulation controlled for the possibility that the monkeys were not actually reporting the number of heard streams but, instead, reported two streams whenever they heard a stimulus that contained high frequencies.

Next, we describe the details of these manipulations. First, on a trial-by-trial basis, we randomly varied the frequency difference between tones A and B. During ~93% of these trials, we presented those frequency differences that provided consistent feedback (i.e., ≤1.0 or ≥10 semitones). For the remaining trials (~7% of the trials or ~44 trials/day), we presented those frequency differences that did not provide consistent feedback (i.e., >1 and <10 semitones).

22

Second, on a trial-by-trial basis, we randomly varied tone A's frequency (range: 865–2226 Hz; mean: 1500 Hz). Third, on a trial-by-trial basis, we varied the sequence duration (i.e., 'listening duration'; 180–2022 ms; mean: 778 ms). Fourth, on a subset of days, we manipulated the temporal relationship between tones A and B. On most days, tones A and B were presented in their standard asynchronous format; see Figure 2-1. However, on select days, tones A and B were presented simultaneously on a randomly subset of trials (~27%); see Figure 2-1 inset. The time between the onsets of the simultaneous was 13 Hz, the same as the asynchronous timing. When tones A and B were presented simultaneously, their frequency difference was always 10 semitones. For the simultaneous trials, the monkeys received rewards independent of their behavioral response.

2.3.4   Data Analyses

We quantified the monkeys' performance by calculating the probability of the monkey reporting two streams (i.e., the monkey moved the joystick to the left). This analysis was conducted as a function of the (a) the frequency difference between tones A and B, (b) the frequency of tone A, (c) listening duration, and (d) the temporal relationship between tones A and B. The 95%-confidence interval on each of these probability values was calculated using the following formula: $1.96*(p*(1-p)/n)^{0.5}$ (Zar, 1996). p was the probability (i.e., the proportion of trials when the monkey reported two streams), and n was the number of trials. The monkeys' performance was considered reliable when the 95%-confidence interval did not overlap with chance performance (i.e., 0.5). A Wilcoxon test was also used to determine whether a probability value differed from chance; the *p*-values that are reported in the text reflect the results of this test. Probability values that were generated from different stimulus-parameter manipulations (e.g., for

23

the upper half of listening durations and the lower half of listening durations) were considered to be significantly ($p<0.05$) different when the 95%-confidence intervals for the two conditions did not overlap.

In a second set of analyses, we conducted two different bootstrap procedures. These bootstrap procedures were conducted to establish performance thresholds, which were then used to identify runs of trials that exceeded these thresholds. The first bootstrap procedure generated a 'null' distribution. This null distribution reflected the probability that the monkeys responded randomly: that is, their responses were independent of the stimulus. To generate this distribution, we first identified those trials in which the frequency difference between tones A and B was 0.5, 1, 10, or 12 semitones and then shuffled the relationship between these frequency differences and the monkeys' reports. Since these frequency differences generate consistent reports in human listeners (Bregman, 1990; Cusack, 2005; Elhilali et al., 2009; Micheyl et al., 2007), when we shuffled the relationship, we hypothesized that we could systematically divorce the stimulus from the response. In contrast, because other frequency differences (i.e., 3 and 5 semitones) do not generate consistent reports in human listeners, there is no 'incorrect' answer and the stimulus cannot be divorced from the response. Therefore, we did not include these trials within our shuffling procedure. Next, we selected, with replacement, N of these shuffled stimulus-report pairings; N was the number of trials/day. We then determined whether a shuffled pair was 'correct' (e.g., frequency difference was ≤1 semitones and the report was 'one stream') or 'incorrect' (e.g., the frequency difference was ≤1 semitones and the report was 'two streams'). Third, to simulate the temporal dynamics of a behavioral session, we treated these shuffled pairs as if they consecutive trials of a behavioral testing session. We then analyzed performance as a function of different running-average window sizes (i.e., 10, 20 or 50 consecutive shuffled

stimulus-response pairings). This procedure was repeated 1000 times for each behavioral session. From this procedure, we generated, as a function of each window size, a distribution of running averages. Finally, we calculated the 'running-average window (RAW) threshold'. In one variant, we calculated the RAW threshold from each session's running-average distribution: the RAW threshold was defined as the upper boundary of each distribution's 95% confidence interval. In a second variant, all of the individual session distributions were pooled together (as a function of window size), and the 'population' RAW threshold was defined as the upper boundary of this pooled distribution's 95% confidence interval.

The second bootstrap procedure generated a distribution of simulated data that, unlike the first bootstrap procedure, maintained the relationship between the auditory stimulus and the monkeys' responses. This bootstrap procedure tested whether, within an experimental session(s), there were temporal epochs or 'runs' of performance that were above chance. First, for each experimental session, we identified those trials in which the frequency difference between tones A and B was 0.5, 1, 10, or 12 semitones; analogous to the logic described above, we did not use the other frequency-difference values. Next, while maintaining the relationship between the stimuli and response, we shuffled the order of the trials. This procedure maintained the relationship between the stimulus and response but disrupted the temporal order of these stimulus-response pairings. Finally, to simulate the temporal dynamics of a behavioral session, we analyzed performance as a function of different running-average window sizes (i.e., 10, 20 or 50 consecutive shuffled stimulus-response pairings). This procedure was repeated 1000 times for each behavioral session. Like with the first bootstrap procedure, we calculated the RAW threshold using the session-by-session-data or the pooled data.

To compare the monkeys' performance with the bootstrapped performance, we extracted

consecutive blocks of data that contained $10, 20,$ or 50 trials in which the frequency difference

was $0.5, 1, 10,$ or 12 semitones. However, because the actual dataset contained trials from all of

the tested-frequency differences, the actual length of the data block could be longer than the

window size. For example, if the window size was 20 trials, the data block might contain 25

trials: 20 trials in which the frequency difference was $0.5, 1, 10,$ or 12 semitones and 5 trials in

which the frequency difference was 3 or 5 semitones. When the monkey's performance on the

$0.5, 1, 10,$ and 12 semitone trials exceeded the RAW threshold, the entire trial block (including

trials in which the frequency difference was 3 or 5 semitones) was considered 'suprathreshold'.

To be clear, the determination of 'suprathreshold' was only based on the 0.5-, 1-, 10-, and 12-

semitone trials because only these trial types were used in the bootstrap procedure. Using the

suprathreshold data, we calculated, as a function of each window size and each frequency

difference, the probability that the monkey reported two streams. These values were generated

from individual behavioral sessions or from the dataset that was generated when the individual

sessions were pooled together, analogous to that done with the bootstrap procedures. Finally, this

analysis was conducted independently for each of the RAW thresholds that were calculated from

each of the two bootstrap procedures (see Table 2-1 for percentage of trials that exceed the RAW

thresholds).

2.4   RESULTS

2.4.1   Monkeys' reports are modulated by the frequency difference between tones A and B

The results from 388 behavioral sessions are shown in Figure 2-3; because monkeys S

and H had comparable behavior, we pooled their behavioral data. Figure 2-3 plots the probability

(i.e., the proportion of trials) that the monkeys reported two auditory streams as a function of

frequency difference between tones A and B. When the frequency difference was ≤1 semitone, the probability that the monkeys reported two streams was less than chance. That is, the probability plus/minus its 95%-confidence interval was less than and did not include 0.5 (i.e., chance performance): 0.5-semitone difference: p=0.454±0.005, $p<0.05$; 1-semitone difference: p=0.465±0.006, $p<0.05$. The interpretation of this result is that the monkeys reliably reported one stream. When the frequency difference was ≥10 semitones, a different pattern emerged: the probability that the monkeys reported two streams exceeded chance: 10-semitone difference: p=0.550±0.007, $p<0.05$; 12-semitone difference: p=0.551±0.005, $p<0.05$. The monkeys' reports for the intermediate frequency differences (3 and 5 semitones) were between the reports for the other frequency differences; however, only the 5-semitone difference did not differ from chance (3-semitone difference: p=0.464±0.018, $p<0.05$; 5-semitone difference: p=0.487±0.020, $p>0.05$).

Although our behavioral data were reliable, the monkeys' behavior clearly did not differ substantially from 0.5 and was poor relative to human performance (Bregman, 1990; Cusack, 2005; Micheyl et al., 2007). However, during the behavioral sessions, we observed short periods (i.e., 10-50 consecutive trials) of high performance. To gain further insight into this observation, we conducted further analyses of their behavior using two different bootstrap procedures.

In the first bootstrap procedure, we shuffled the relationship between the auditory stimulus and the monkeys' responses to generate a null distribution. This distribution tested the hypothesis that, over short windows of trials, the monkeys performed better than chance and were using the stimulus to guide their choices. Panels A and B in Figure 2-4 show the RAW thresholds that were generated from this procedure and the respective suprathreshold subset of behavioral data (see **Methods**). Figure 2-4A shows the monkeys' performance when the RAW thresholds were calculated from the population data. This threshold calculation is a reflection of

performance for a given running average relative to the monkeys' general behavior. Figure 2-4B shows the monkey's performance using the session-by-session RAW thresholds. These thresholds provide a measure of performance relative to a particular day's behavior. We again found that monkeys (a) significantly reported one stream at the smallest frequency differences; (b) significantly reported two streams at the largest frequency differences; and (c) for intermediate frequency differences, behavior did not differ from chance (i.e., it fell below the running-average threshold). More specifically, for running windows of 10 trials (green data), we found that the monkeys' behavior was ~30% better than their overall behavior that was shown in Figure 2-3. The monkeys' performance improved modestly for larger running-average windows (blue and red data): for windows of 50 trials (red data), behavior improved by ~10%. Like the data in Figure 2-3, this bootstrap analysis indicated that the monkeys' behavior was guided by the stimuli. However, unlike the data shown in Figure 2-3, this bootstrap analysis indicated that—under certain circumstances—the monkeys' performance can closely approximate the performance of human listeners.

To further evaluate these windows of high performance, we performed a second bootstrap procedure. In this procedure (and unlike the first one), we maintained the integrity of the stimulus-response pairings but shuffled the temporal order of these pairings. This procedure tested explicitly the reliability of the running-average windows; that is, this procedure tested whether there were short 'runs' of performance that were above chance. Figures 2-4C and 2-4D show the monkeys' performance for those runs of trials that exceeded the bootstrap's performance at each of the RAW thresholds. Once again, we identified runs of trials in which the monkeys' behavior exceeded the RAW thresholds. We again found that short running-average windows of 10 trials (green data) were ~30% than the overall data in Figure 2-3; with more

modest gains of ~10% over the overall data for windows of 50 trials (red data).

Together, all three analyses indicate that the monkeys successfully learned the streaming task. Using all of the data (Fig. 2-3), we found that their performance was reliable, and the pattern of their behavior was consistent—albeit poorer—than human performance. However, importantly, we found periods of high performance, defined as having a running average that fell above the RAW thresholds. These periods of high performance, which more closely approximated human performance, were found in windows of 10-50 trials (Fig. 2-4).

2.4.2   The monkeys' behavior was independent of tone A's frequency

Next, we tested whether the trial-by-trial variability in the frequency of tone A (range: 865-2226 Hz) affected the monkeys' behavioral reports. As a reminder, because the frequency of tone B was based on tone A's frequency, when we changed tone A's frequency, we changed the frequency content of the auditory sequence. This analysis is important because if the monkeys were using a strategy of reporting 'two streams' whenever they heard a high-frequency stimulus, then changing the frequency of tone A should affect their behavior. However, if the monkeys were simply reporting the number of heard streams, their reports should be independent of tone A's frequency. The results of this analysis are shown in Figure 2-5. In this Figure, we again plot the probability that the monkeys' reported two streams as a function of the frequency difference between tones A and B. However, here, we subdivided the data: the 'low-frequency' data contained the monkeys' reports when the tone A's frequency was between 865-1500 Hz (the lower half of the distribution of tone A frequencies), whereas the 'high-frequency' data contained reports when tone A's frequency was 1501-2226 Hz (the upper half of the distribution of tone A frequencies). Using the two bootstrap procedures (see **Methods**), we calculated the running-

average thresholds independently for both the low-frequency and high-frequency data groups; because data for all RAW thresholds followed the same pattern, Figure 2-5 only shows the data relative to the 20-trial RAW threshold. As can be seen, for each of those frequency differences that exceeded the bootstrap threshold (i.e., 0.5, 1, 10 and 12 semitones), in most cases, the confidence intervals on the monkeys' reports for the low-frequency data overlapped with those of the high-frequency data. That is, the frequency of tone A did not significantly (p>0.05) affect the monkeys' reports. When the confidence intervals did not overlap, we could not identify any consistent trend between the frequency of tone A and the monkeys' reports. These results are consistent with the hypothesis that the monkeys' reports were independent of tone A's frequency.

2.4.3   Longer stimulus durations biased the monkeys to report two streams

Next, we tested how the trial-by-trial variability in the amount of that the monkeys' listened to the auditory sequence time (listening duration; 180-2022 ms) affected their behavior. We divided the behavioral into trials when the listening duration was 180-770 ms (the lower half of the distribution of listening durations) and into trials when the listening duration was 771-2022 ms (the upper half of the distribution of listening durations). The results of this analysis are shown in Figure 2-6; because data for all RAW thresholds followed the same pattern, Figure 2-6 only shows the data relative to the 20-trial RAW threshold. As can be seen, for each of those frequency differences that exceeded the bootstrap threshold (i.e., 0.5, 1, 10 and 12 semitones), the confidence intervals on the monkeys' reports for the longer-duration data never overlap with, and are always higher than, those of the shorter-duration sequences. Like human listeners (Micheyl et al., 2007), longer-duration sequences biased the monkeys to report 'two streams' more often than shorter-duration sequences.

2.4.4 Simultaneous presentation of tones A and B biases the monkeys to report one stream

Finally, we tested whether the temporal relationship of tone A and tone B affected the monkeys' behavioral reports. If, as discussed above, the monkeys were simply reporting 'two streams' whenever they perceived a high-frequency stimulus, their reports should not depend on the tones A and Bs' temporal relationship. However, if the monkeys were reporting the number of heard streams, then, like human listeners (Elhilali et al., 2009), their reports should be biased toward reporting one stream when tone A and B were presented simultaneously and even when the frequency difference between tones A and B is large (e.g., $\geq$ 10 semitones).

Because the simultaneous presentation of tones A and B sounded different than the normal asynchronous presentation, we limited its presentation to a small subset of behavioral sessions (N = 18). Consequently, this data set was not large enough for our bootstrap procedure. Finally, to maximize the informative trials with the least exposure to the simultaneous trials as possible, we limited this presentation to a 10-semitone frequency difference.

Figure 2-7 shows the results of this analysis. As noted above, when the tones were asynchronous and the frequency difference was 10 semitones, the probability that the monkeys reported two streams was significant (p=0.524±0.007; $p<0.05$; this proportion represents the monkeys' behavior during those sessions when simultaneous tones were also presented). However, when tones A and B were presented simultaneously, the monkeys were more likely to report one stream (10 frequency semitones: p=0.459±0.051). This proportion of trials was significantly ($p<0.05$) smaller than the one when tones A and B were presented asynchronously. However, it is not different than chance performance (0.5; $p>0.05$). Nonetheless, this result is consistent with the hypothesis that the simultaneous presentation of tones A and B biased the

monkeys toward reports of 'one stream'.

2.5   DISCUSSION

The streaming task has been used extensively to test auditory perception in humans. Here, we demonstrated for the first time that rhesus macaques' behavioral reports were consistent with those of human listeners. We found that monkeys reported small frequency differences as one stream, large ones as two streams, and intermediate ones as either one or two streams. We further found that the monkeys' reports were independent of the absolute frequency content of the stimulus but that longer listening durations biased the monkeys toward reporting two streams. Moreover, simultaneous presentation of tones A and B biased the monkey toward reporting one stream. Below, we discuss the interpretation of our findings, as well as caveats regarding performance and implications for auditory processing across species.

Although our current findings are consistent with human studies, training monkeys on the streaming task presented challenges that are not faced in training humans on this task. Namely, monkeys could not be explicitly told to report one or two streams. Therefore, without controls, our results could have been interpreted as the monkeys merely reporting any stimulus with a high frequency as two streams and anything else as one stream. However, three controls support the hypothesis that the monkeys were reporting the number of heard streams. First, by presenting tone A across a range of frequencies that spanned nearly 2.5 octaves—considerably larger than the frequency difference between tones A and B—we demonstrated that the monkeys' reports were independent of the frequency of tone A (Fig. 2-5). Second, like human listeners (Micheyl et al., 2007), longer stimulus durations biased the monkeys to report two streams. This result is consistent with findings that the perception of two streams 'builds up' over time (Elhilali et al., 2009; Micheyl et al., 2007) and is inconsistent with a hypothesis of simply reporting frequency

32

differences. Finally, similar to human listeners (Elhilali et al., 2009), when the tone bursts were presented simultaneously and the frequency difference was large (which normally elicits reports of 'two streams), the monkeys' reports were biased toward those of 'one stream' (Fig. 2-7). Overall, these controls are consistent with the hypothesis that the monkeys reported the number of heard streams.

Simultaneously presenting tones was a particularly important control because it showed that the monkeys were actually reporting their streaming percept instead of merely reporting whether or not they heard a high frequency tone. When tones were presented asynchronously, monkeys might have used a strategy in which they categorized whether or not a high frequency tone was present. However, because the synchronously presented chord has a high frequency tone but monkeys were biased towards reporting one stream, a frequency-content categorization cannot wholly explain their performance. Furthermore, it should be noted that all of our frequency separations were distinguishable by rhesus macaques, and therefore, monkeys should be able to distinguish the tones in each trial (Sinnott et al., 1985) (and therefore, monkeys must based their decision on stream percept, not by categorizing whether they heard one repeated frequency, or two alternating frequencies).

Although the monkeys' performance was reliable and the three stimulus controls yielded results qualitatively similar to those of humans, the monkeys overall performance (Fig. 2-3) indicated that this task was difficult. However, in observing the monkeys' performance, it was apparent that there were times when the monkeys had short runs of good performance. Indeed, our two bootstrap procedures indicated that the monkeys used the stimulus to guide their behavior and had high levels of performance over windows of 10-50 trials (Fig. 2-4) that more closely mirrored that of human-performance levels (Cusack, 2005; Elhilali et al., 2009; Micheyl et al.,

2007). Importantly, since trials with a given frequency difference were randomly distributed within a session, these periods of high performance did not represent runs of 'easy' trials (e.g., blocks when the same frequency difference was presented multiple times in succession).

How do our results fit into the general comparative psychophysical literature? Our findings support this literature, much of which has found that humans and non-human animals similarly process auditory stimuli. For example, several sets of studies have found that humans, monkeys, quail and chinchillas have similar categorical boundaries for human phonemes (Kuhl et al., 1975b; Kuhl et al., 1982). Similarly, monkeys exhibit amodal completion in a manner similar to humans (Miller et al., 2001; Petkov et al., 2003; Petkov et al., 2007) and group sounds in a manner similar to humans (Izumi, 2002). Other studies have demonstrated that non-human animals parse the auditory scene like human listeners (Aulanko et al., 1993; Coath et al., 2005; DeWitt et al., 2012; Narayan et al., 2007; Noda et al., 2012). Finally, our data are consistent with those studies that used indirect assays of streaming (Izumi, 2002; Ma et al., 2010; Noda et al., 2012).

Where in the brain is this information being processed? Several studies have recorded from the monkey primary auditory cortex while monkeys were listening passively to auditory sequences similar to those used in our study (Fishman et al., 2004; Fishman et al., 2001a; Micheyl et al., 2005). Although the monkeys were not actively engaged in a streaming task during these studies, the pattern of neural activity indicated that this cortical region may be involved in the grouping and segregation of auditory stimuli into auditory streams. Indeed, other sets of findings in the core and belt regions of the auditory cortex have also hinted at a role for these brain regions in auditory scene analysis (Bendor et al., 2006; Fishman et al., 2004; Fishman et al., 2000; Fishman et al., 2001a; Fishman et al., 2001b; Micheyl et al., 2007; Niwa et al., 2012b; Tomasello,

34

2008; Wang et al., 2008). More generally, the ventral auditory pathway, which is specialized for mediating auditory perception (Bizley et al., 2013a; Cohen, 2012; Kaas et al., 1999; Rauschecker et al., 2009; Romanski et al., 2009), likely plays a role in the neural computations that allow a listener to segregate or group an auditory stimulus into one or more auditory streams.

Finally, this task will provide a powerful tool to disassociate brain activity that is related to the features of the auditory stimulus from activity that is related to a listeners' behavioral report. In particular, since listeners reports vary, on a trial-by-trial basis, for sequences with intermediate frequency differences (>1 semitone and <10 semitones), this stimulus can be considered akin to a 'bistable percept' (Andersen et al., 1996; Bregman, 1990; Logothetis et al., 1989; Parker et al., 1998). In other words, by holding the stimulus constant and analyzing neural responses as a function of the listener's behavioral report, we can identify and differentiate between the brain regions and the computations that underlie auditory scene analysis, auditory perception and decision-making.

2.6   CONCLUSION

In conclusion, we have shown that monkeys can be trained to perform the streaming task. Moreover, their behavioral reports are consistent with human reports across a variety of experimental manipulations. These findings add further evidence that monkeys group and segregate acoustic stimuli similarly to humans. Therefore, they provide an excellent model to study the neural coding that underlies this behavior, and more generally, auditory perception.
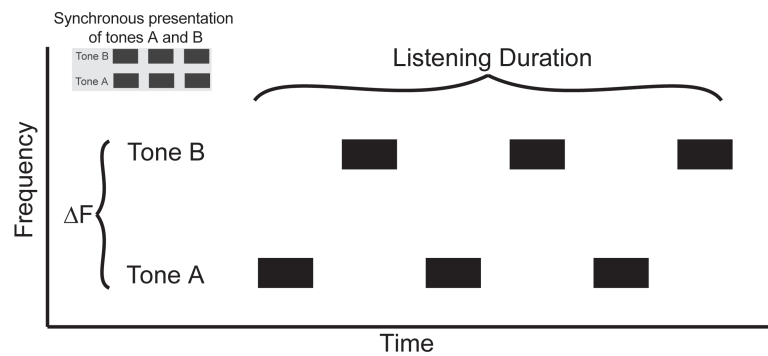
Figure 2-1: Schematic of the auditory stimulus to test auditory streaming.
The auditory stimulus was an asynchronous sequence of two types of tone bursts: tone A and tone B. Typically, tones A and B were presented asynchronously but were at times presented simultaneously (see inset at upper left). The frequency of tone A, the frequency difference between the tones A and B (ΔF), and the listening duration (i.e., the duration of the auditory sequence) varied on a trial-by-trial basis. The units on the x- and y-axes are arbitrary.
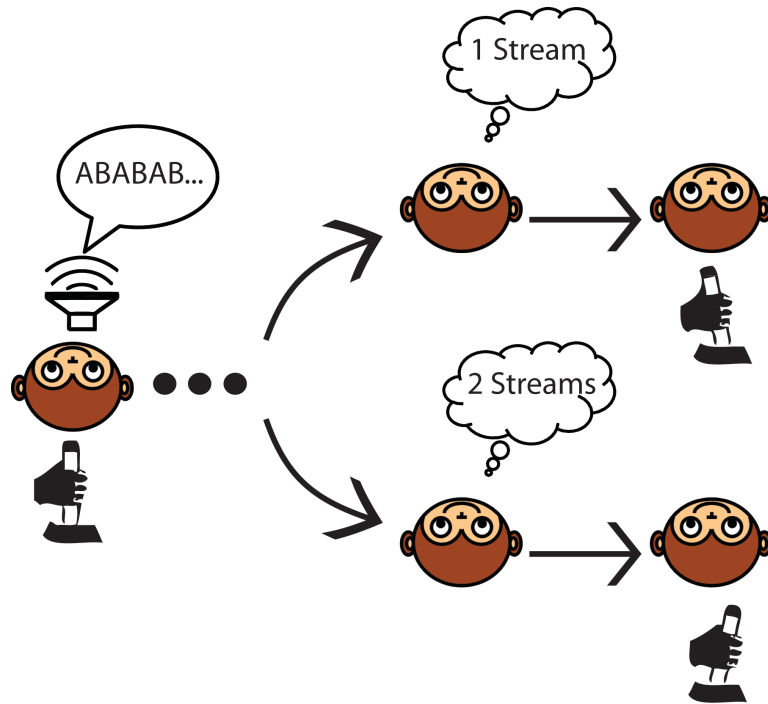
Figure 2-2: Schematic of the streaming task.
The streaming task is a one-interval, two-alternative, forced-choice task requiring a monkey to report whether he heard one or two auditory streams by moving a joystick to the right (one stream) or left (two streams). When the frequency difference between tones A and B was ≤1 semitone or ≥10 semitones, the monkeys received a juice reward for reporting the correct answer. For all other frequency differences, the monkeys received a reward on 50% of randomly selected trials; the decision to reward was made independent of their behavioral report.
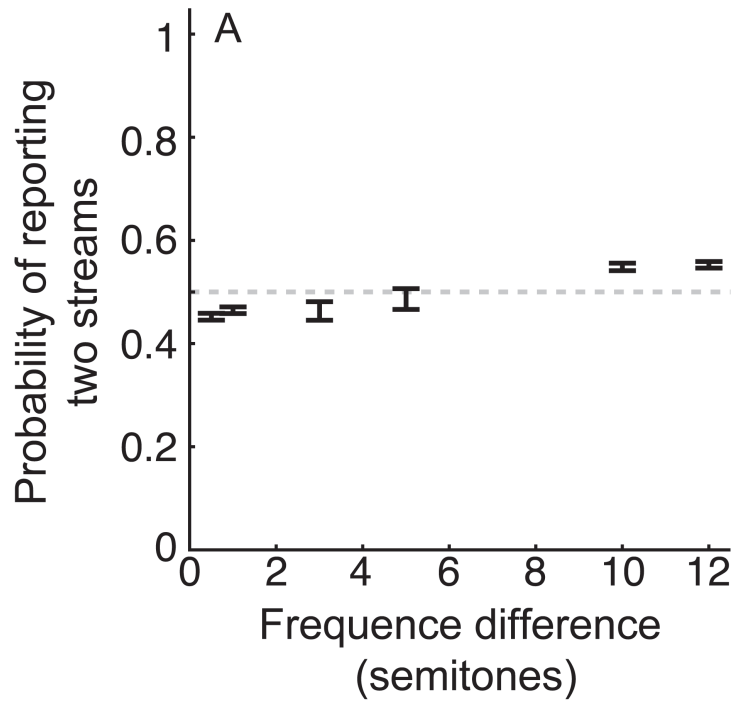
Figure 2-3:Behavioral performance: all data and all sessions.
The average performance of both monkeys from all of the behavioral sessions reported in this manuscript (except for those trials when tone A and B were presented simultaneously; see Fig. 2-7). The center of each bar indicates the average probability (i.e., the proportion of trials) that the monkeys reported two streams; the length of the bars indicates the 95% confidence interval. The gray dashed line represents chance performance (0.5) of answering one or two streams.
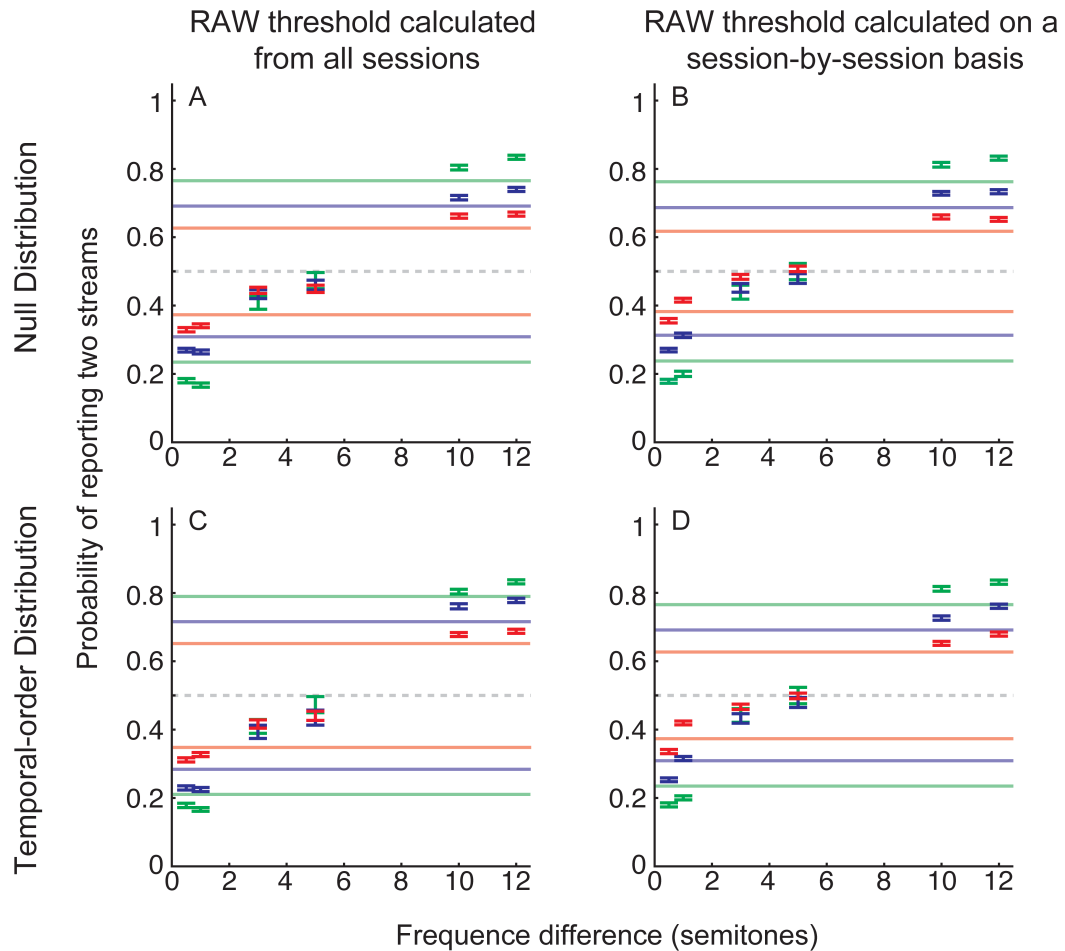
Figure 2-4: Behavioral performance: behavior relative to the bootstrapped RAW thresholds.

The data on the top row show the monkeys' behavior relative to a bootstrapped null distribution (i.e., one in which there is no relationship between the stimulus and the monkeys' responses). The data on the bottom row show the monkeys' behavior relative a second bootstrap distribution that maintained the integrity between the stimulus and the monkeys' responses but shuffled the temporal order. This bootstrap procedure tested explicitly whether there were significant temporal runs of performance. For data in the left column, the RAW thresholds were calculated from data that was pooled across all behavioral sessions. For data in the right column, the RAW thresholds were calculated on a session-by-session basis. The color of each of the solid lines illustrates the upper and lower boundaries of the different RAW thresholds: green is 10 trials, blue is 20 trials, and red is 50 trials. The center of each bar indicates average suprathreshold performance; the color of the data points is consistent with the color of the threshold values. The length of the bars indicates the 95% confidence interval. If error bars from one color are not visible, it is because the confidence intervals for multiple conditions overlap completely. The gray dashed line represents chance performance (0.5) of answering one or two streams.

RAW threshold calculated from all sessions

RAW thresholdcalculated on a session-by-session basis
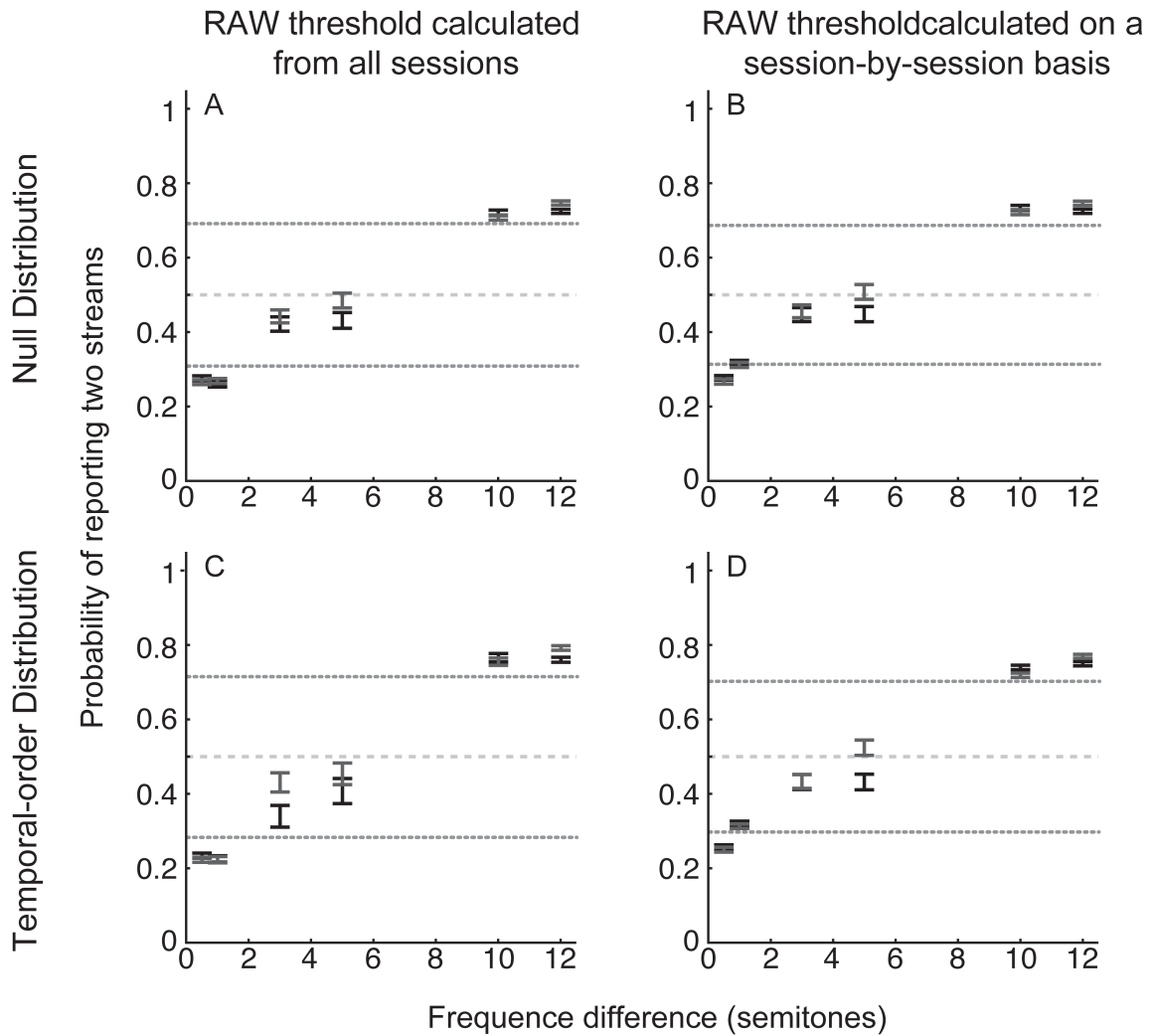
Figure 2-5: Behavioral performance: dependence on the frequency of tone A.
The data in each row and column are organized analogous to that in Figure 2-4. The dotted lines illustrate the upper and lower boundaries of the 20-trial RAW threshold; the other thresholds are not shown. The data in black indicate average suprathreshold performance when the frequency of tone A was relatively low (865-1500 Hz). The data in gray indicate average suprathreshold performance when the frequency of tone A was relatively high (1501-2226 Hz). The center of each bar indicates average suprathreshold performance; the length of the bars indicates the 95% confidence interval. If error bars from one color are not visible, it is because the confidence intervals for multiple conditions overlap completely. The gray dashed line represents chance performance (0.5) of answering one or two streams.
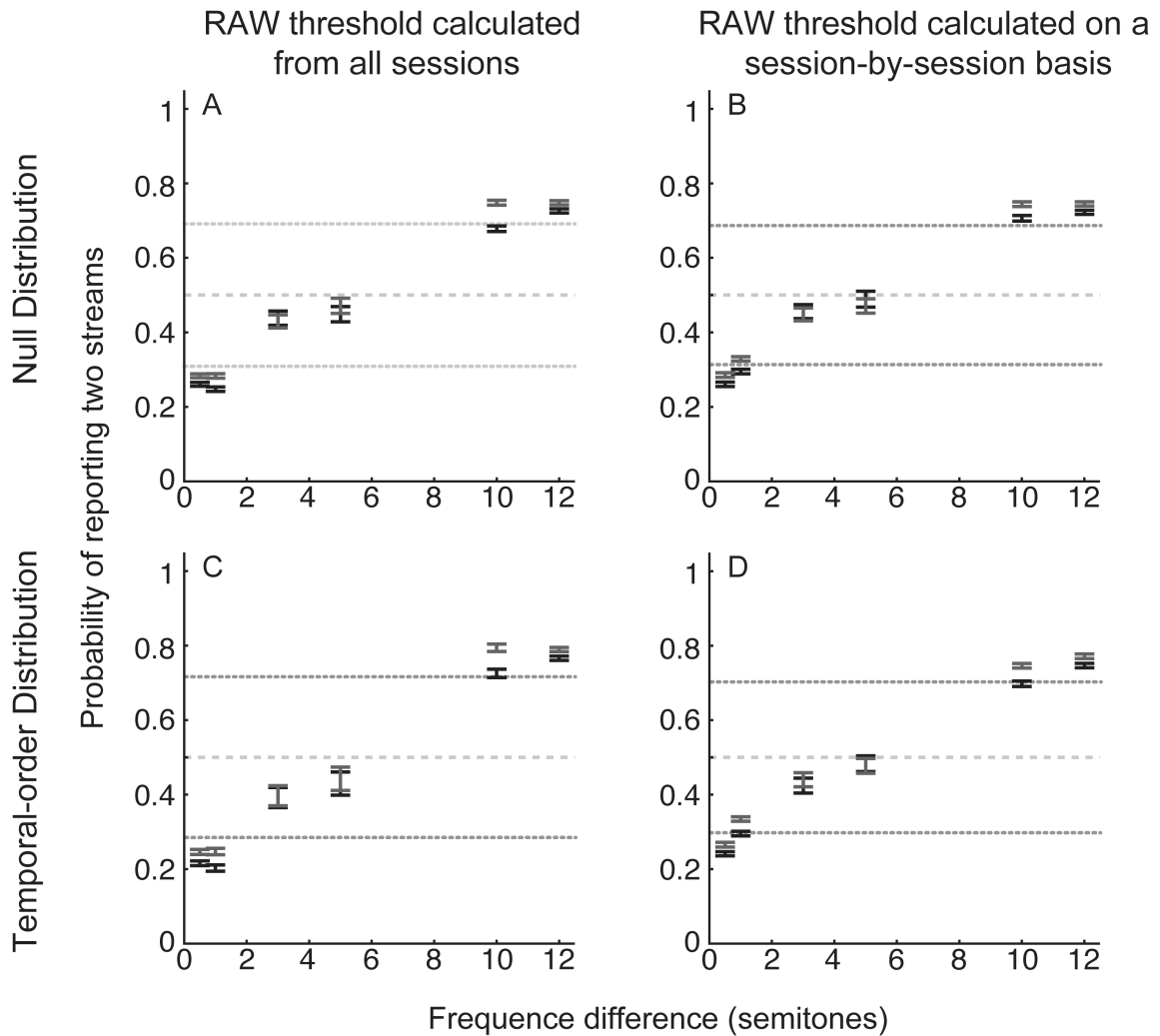
Figure 2-6: **Behavioral performance: dependence on listening duration.**
The data in each row and column are organized analogous to that in Figure 2-4. The dotted lines illustrate the upper and lower boundaries of the 20-trial RAW threshold; the other thresholds are not shown. The data in black indicate average suprathreshold performance when the listening duration was short (180-770 ms). The data in gray indicate average suprathreshold performance when the listening duration was long (771-2022 ms). The center of each bar indicates average suprathreshold performance; the length of the bars indicates the 95% confidence interval. If error bars from one color are not visible, it is because the confidence intervals for multiple conditions overlap completely. The gray dashed line represents chance performance (0.5) of answering one or two streams.
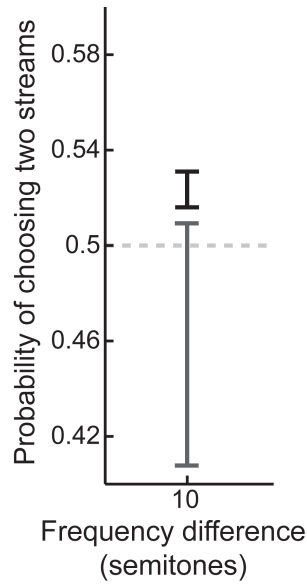
Figure 2-7: Behavioral performance: dependence on the temporal structure of tones A and B.
The black bar indicates average performance for trials when tones A and B were presented asynchronously. The gray bar indicates average performance for trials when tones A and B were presented simultaneously. The center of each bar indicates the average probability (i.e., the proportion of trials) that the monkeys reported two streams; the length of the bars indicates the 95% confidence interval.

Bootstrap Type

		Null (Population)	Null (By session)	Temporal (Population)	Temporal (By session)
	10	.08	.08	.08	.07
	20	.1	.11	.05	.06
	50	.09	.12	.06	.07

Trials in RAW

Table 2-1: Monkeys' performance exceeds RAW threshold for a reliable proportion of trials
The table summarizes the proportion of trials in which the monkeys' performance exceeded the RAW threshold. For all bootstrap procedures and RAW thresholds, the monkeys performance exceeded chance.

2.9 REFERENCES

Andersen, R.A., Bradley, D.C., Shenoy, K.V. 1996. Neural mechanisms for heading and structure-from-motion perception. Cold Spring Harb Symp Quant Biol 61, 15-25.

Aulanko, R., Hari, R., Lounasmaa, O.V., Naatanen, R., Sams, M. 1993. Phonetic invariance in the human auditory cortex. Neuroreport 4, 1356-8.

Bendor, D., Wang, X. 2006. Cortical representations of pitch in monkeys and humans. Curr Opin Neurobiol 16, 391-9.

Bizley, J.K., Cohen, Y.E. 2013a. The what, where and how of auditory-object perception. Nat Rev Neurosci 14, 693-707.

Bregman, A.S. 1990. Auditory Scene Analysis: The Perceptual Organization of Sound MIT Press, Cambridge, MA.

Bregman, A.S., Ahad, P.A., Crum, P.A.C., O'Reilly, J. 2000. Effects of time intervals and tone durations on auditory stream segregation. Percept Psychophys 62, 626-636.

Carlyon, R.P., Cusack, R., Foxton, J.M., Robertson, I.H. 2001. Effects of Attention and Unilateral Neglect on Auditory Stream Segregation. J Exp Psychol Hum Percept Perform 27, 115-127.

Coath, M., Brader, J.M., Fusi, S., Denham, S.L. 2005. Multiple views of the response of an ensemble of spectro-temporal features support concurrent classification of utterance, prosody, sex and speaker identity. Network 16, 285-300.

Cohen, Y.E. 2012. Auditory Cognition: The Integration of Psychophysics with Neurophysiology. In: Cohen, Y.E., Popper, A.N., Fay, R.R., (Eds.), Neural Correlates of Auditory Cognition, Vol. Springer Handbook of Auditory Research. Springer-Verlag, New York. pp. 1-6.

Cusack, R. 2005. The Intraparietal Sulcus and Perceptual Organization. J Cogn Neurosci 17, 641-651.

DeWitt, I., Rauschecker, J.P. 2012. Phoneme and word recognition in the auditory ventral stream. Proc Natl Acad Sci U S A 109, E505-14.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A.J., Shamma, S.A. 2009. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. Neuron 61, 317-29.

Fishman, Y.I., Arezzo, J.C., Steinschneider, M. 2004. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J Acoust Soc Am 116, 1656-70.

Fishman, Y.I., Reser, D.H., Arezzo, J.C., Steinschneider, M. 2000. Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness. J Acoust Soc Am 108, 235-46.

Fishman, Y.I., Reser, D.H., Arezzo, J.C., Steinschneider, M. 2001a. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear Res 151, 167-187.

Fishman, Y.I., Volkov, I.O., Noh, M.D., Garell, P.C., Bakken, H., Arezzo, J.C., Howard, M.A., Steinschneider, M. 2001b. Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. J Neurophysiol 86, 2761-88.

Griffiths, T.D., Warren, J.D. 2004. What is an auditory object? Nat Rev Neurosci 5, 887-892.

Horvath, J., Czigler, I., Sussman, E., Winkler, I. 2001. Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. Cognitive Brain Research 12, 131-144.

Izumi, A. 2002. Auditory stream segregation in Japanese monkeys. Cognition 82, B113-B122.

Kaas, J.H., Hackett, T.A. 1999. 'What' and 'where' processing in auditory cortex. Nat Neurosci 2, 1045-1047.

Kuhl, P.K., Miller, J.D. 1975b. SPEECH-PERCEPTION BY CHINCHILLA - PHONETIC BOUNDARIES FOR SYNTHETIC VOT STIMULI. J. Acoust. Soc. Am. 57, S49-S50.

Kuhl, P.K., Padden, D.M. 1982. ENHANCED DISCRIMINABILITY AT THE PHONETIC BOUNDARIES FOR THE VOICING FEATURE IN MACAQUES. Perception & Psychophysics 32, 542-550.

Logothetis, N.K., Schall, J.D. 1989. Neuronal correlates of subjective visual perception. Science 245, 761-3.

Ma, L., Micheyl, C., Yin, P., Oxenham, A.J., Shamma, S.A. 2010. Behavioral measures of auditory streaming in ferrets (Mustela putorius). J Comp Psychol 124, 317-30.

McDermott, J. 2009. The cocktail party problem. Curr Biol 19, R1024-R1027.

Micheyl, C., Tian, B., Carlyon, R.P., Rauschecker, J.P. 2005. Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron 48, 139-48.

Micheyl, C., Carlyon, R.P., Gutschalk, A., Melcher, J.R., Oxenham, A.J., Rauschecker, J.P., Tian, B., Courtenay Wilson, E. 2007. The role of auditory cortex in the formation of auditory streams. Hear Res 229, 116-31.

Miller, C.T., Dibble, E., Hauser, M.D. 2001. Amodal completion of acoustic signals by a nonhuman primate. Nat Neurosci 4, 783-4.

Narayan, R., Best, V., Ozmeral, E., McClaine, E., Dent, M., Shinn-Cunningham, B., Sen, K. 2007. Cortical interference effects in the cocktail party problem. Nat Neurosci 10, 1601-7.

Niwa, M., Johnson, J.S., O'Connor, K.N., Sutter, M.L. 2012b. Activity related to perceptual judgment and action in primary auditory cortex. J Neurosci 32, 3193-210.

Noda, T., Kanzaki, R., Takahashi, H. 2012. Oscillatory neuronal synchronization for auditory stream segregation in auditory cortex of unanesthetized rat. In: Murray, M., (Ed.), 4th Conference on Auditory Cortex, Lausanne, Switzerland. pp. A20.

Parker, A.J., Newsome, W.T. 1998. Sense and the single neuron: probing the physiology of perception. Annu Rev Neurosci 21, 227-77.

Petkov, C.I., O'Connor, K.N., Sutter, M.L. 2003. Illusory sound perception in macaque monkeys. J Neurosci 23, 9155-61.

Petkov, C.I., O'Connor, K.N., Sutter, M.L. 2007. Encoding of illusory continuity in primary auditory cortex. Neuron 54, 153-65.

Rahne, T., Sussman, E. 2009. Neural representations of auditory input accommodate to the context in a dynamically changing acoustic environment. European Journal of Neuroscience 29, 205-211.

Rauschecker, J.P., Scott, S.K. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12, 718-24.

Recanzone, G.H., Sutter, M.L. 2008. The biological basis of audition. Annu Rev Psychol 59, 119-42.

Romanski, L.M., Averbeck, B.B. 2009. The primate cortical auditory system and neural representation of conspecific vocalizations. Annu Rev Neurosci 32, 315-46.

Shamma, S. 2008. On the Emergence and Awareness of Auditory Objects. PLoS Biol 6, e155.

Shamma, S., Micheyl, C. 2010. Beyond the scenes of auditory perception. Curr Opin Neurobiol 20, 1-6.

Shamma, S.A., Elhilali, M., Micheyl, C. 2011. Temporal coherence and attention in auditory

scene analysis. Trends Neurosci 34, 114-23.

Shinn-Cunningham, B.G. 2008. Object-based auditory and visual attention. Trends Cogn Sci 12, 182-6.

Sinnott, J.M., Petersen, M.R., Hopp, S.L. 1985. Frequency and intensity discrimination in humans and monkeys. J Acoust Soc Am 78, 1977-85.

Sussman, E.S. 2005. Integration and segregation in auditory scene analysis. The Journal of the Acoustical Society of America 117, 1285-1298.

Sussman, E.S., Horvath, J., Winkler, I., Orr, M. 2007. The role of attention in the formation of auditory streams. Perception & Psychophysics 69, 136-152.

Tomasello, M. 2008. Origins of Human Communication MIT Press, Cambridge, MA.

Wang, X., Lu, T., Bendor, D., Bartlett, E. 2008. Neural coding of temporal information in auditory thalamus and cortex. Neuroscience 154, 294-303.

Winkler, I., Denham, S.L., Nelken, I. 2009. Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci 13, 532-40.

Zar, J. 1996. Biostatistical Analysis Prentice Hall, Upper Saddler River, NJ.

3 CHAPTER 3: THE CONTRIBUTION OF PRIMARY AUDITORY CORTEX TO AUDITORY STREAMING

3.1 ABSTRACT

The contribution of the auditory cortex to perception remains controversial. While monkeys reported whether a temporal sequence of tone bursts was heard as one or two auditory streams, we recorded from sites in primary auditory cortex (A1). Like earlier work, A1 had frequency-tuned responses that habituated, independent of frequency content, as the auditory sequence unfolded over time. We report for the first time that A1 firing rate was modulated by the monkeys' choices; this modulation increased with listening time. Thus, A1 activity contributes to the sensory evidence underlying the segregation and grouping of acoustic stimuli into distinct auditory streams. However, because this modulation happens even the absence of frequency-dependent differences in habituation, it puts our data at odds with a prominent hypothesis proposed by Micheyl et al.'s (2005), arguing for frequency-dependent habituation as a coding mechanism for this streaming percept. We propose that task-dependent differences in frequency tuning underlie these different findings.

3.2 INTRODUCTION

Auditory perception is mediated in the ventral auditory pathway (Bizley et al., 2013a; Hackett, 2011; Rauschecker et al., 2009; Romanski et al., 2009). In rhesus monkeys, this pathway begins in core auditory cortex, which includes primary auditory cortex (A1) and area R. Although there is broad agreement that the ventral pathway has a critical role in auditory perception, there is not a consensus on the distinct contributions of different regions of this pathway to perception (Bizley et al., 2013a; Giordano et al., 2012; Rauschecker, 2012). In particular, there remains considerable controversy regarding the contribution of the auditory cortex to perception (Binder

47

et al., 2004; Bizley et al., 2013b; Gutschalk et al., 2005; Lemus et al., 2009; Mesgarani et al., 2012; Niwa et al., 2012a; Niwa et al., 2013; Tsunada et al., 2011a).

To directly address a contribution of A1 to auditory perception, we recorded neural activity in rhesus monkeys while they simultaneously participated in an auditory-streaming task. During this task, which used conditions comparable to those in human studies (Christison-Lagay et al., 2014a), monkeys reported whether a temporal sequence of tone bursts—in which tone bursts alternated between two frequencies—was heard as one auditory stream or two auditory streams; an auditory stream is a single perceptual auditory unit, akin to a visual object (Bizley et al., 2013a; Bregman, 1990). Although with certain combinations of tone-burst frequencies, listeners reliably report one or two auditory streams, for other combinations, their reports vary trial-by-trial, despite the fact that the auditory stimulus is physically identical (Bregman, 1990; Griffiths et al., 2004; McAdams et al., 1979). This is advantageous because it allows a differentiation between neural representations of an acoustic stimulus versus representations of a reported percept.

We found that, like earlier work (Fishman et al., 2004; Fishman et al., 2001a; Micheyl et al., 2005), A1 neurons had frequency-tuned responses that habituated, independent of frequency content, as the auditory sequence unfolded over time. Our study substantially advanced these prior findings by directly identifying a relationship between A1 firing rates and the perceptual reports of the monkey. Specifically, we found that the tone-burst-by-tone-burst firing rate of A1 neurons was modulated by the monkeys' choices, which increased with listening time. These findings provide the first direct evidence that A1 activity can contribute to the sensory evidence underlying the segregation and grouping of acoustic stimuli into distinct auditory streams. However, because this modulation happens even the absence of frequency-dependent differences

in habituation, it puts our data at odds with a prominent hypothesis proposed by Micheyl et al.'s (2005), arguing for frequency-dependent habituation as a coding mechanism for this streaming percept. We propose that task-dependent differences in frequency tuning underlie these different findings.

3.3 EXPERIMENTAL PROCEDURE

The University of Pennsylvania Institutional Animal Care and Use Committee approved all experimental protocols. All surgical procedures were conducted under general anesthesia and using aseptic surgical techniques.

3.3.1 Experimental chamber.

Behavioral training and recording sessions were conducted in an electrically shielded, darkened room with sound-absorbing walls. During each session, a monkey (*Macaca mulatta*; monkey H or monkey S) was seated in a primate chair in the center of the room. A calibrated speaker (model MSP7, Yamaha) was placed in front of the monkey at a distance of 1.5 m and at eye level. A touch-sensitive joystick was attached to the primate chair; the monkey moved the joystick to indicate his behavioral report. All auditory stimuli were generated using the RX6 digital-signal-processing platform (TDT Inc.) and were transduced by the Yamaha speaker.

3.3.2 Targeting of the primary auditory cortex.

From MRI images of each monkey's skull and brain, the stereotactic location of A1 was identified using the Brainsight software package (Rogue Technologies). A1 was located on the surface of the superior temporal gyrus (Fig. 3-3A; monkey H: right hemisphere; monkey S: left hemisphere). A1 was further defined by its neural response properties (Kajikawa et al., 2005;

49

Kajikawa et al., 2011; Kusmierek et al., 2009; Rauschecker et al., 2004; Recanzone et al., 2000).

3.3.3   Auditory tasks and stimuli.

The 'best-frequency' task identified the best frequency of an A1 recording site. The 'streaming' task tested the ability of a monkey to segregate a tone-burst sequence into one or two auditory streams (Christison-Lagay et al., 2014a). Data from the best-frequency task were integrated into the stream task, as described below.

Best-frequency task.

A monkey listened passively while individual tone bursts were presented in a random order. The tone bursts (100-ms duration with a 5-ms $\cos^2$ ramp; 65 dB SPL) varied between 0.4–4 kHz in one-quarter octave steps.

Streaming task.

The streaming task was a single-interval, two-alternative-forced-choice discrimination task that required a monkey to report whether he heard one or two auditory streams. 500 ms after the monkey grasped the joystick, we presented a temporal sequence of tone bursts. Following offset of the auditory sequence, an LED was illuminated, which signaled the monkey to indicate his behavioral report. The monkey moved the joystick to the (1) right to report 'one auditory stream' or (2) left to report 'two auditory streams' (Fig. 3-1A).

Each tone burst (40-ms duration with a 5-ms $\cos^2$ ramp; 65 dB SPL; 13-Hz inter-tone-burst-interval) in the temporal sequence alternated between two frequencies: 'tone A' and 'tone B'; Fig. 3-1B. Tone A was always set to a recording site's best frequency (see *Data-collection strategy* below), whereas tone B was presented either at 0.5, 3, 5, or 12 semitones above this best frequency. The frequency of tone B and the duration (mean: 750±150 ms) of each tone-burst

sequence (i.e., 'listening time') varied on a trial-by-trial basis.

Training procedure and reward schedule.

Initially, a monkey was trained on tone-burst sequences in which tones A and B were separated either by ≤1.0 or ≥10 semitones. These frequency differences were chosen because, in human listeners (Bregman, 1990; Cusack, 2005; Micheyl et al., 2005) and recently confirmed in monkeys (Christison-Lagay et al., 2014a), they reliably elicit reports of one or two streams, respectively. The monkey was only rewarded for correct responses. A response was correct when (1) the frequency difference between tone A and tone B was ≤1.0 semitone, and the monkey moved the joystick to the right; or (2) the frequency difference was ≥10 semitones, and he moved the joystick to the left (Fig. 3-1A).

Following the stabilization of a monkey's performance with these tone-burst sequences, we presented sequences that contained 'intermediate' frequency differences (i.e., 3 and 5 semitones). These sequences do not elicit reliable reports of either one or two streams in human or monkey listeners (Bregman, 1990; Christison-Lagay et al., 2014a; Cusack, 2005; Elhilali et al., 2009; Micheyl et al., 2007). Because on these trials, there was not a 'correct' answer, we rewarded the monkey on 50% of randomly selected trials, independent of his behavioral report.

3.3.4   Neural-recording methodology.

Prior to a recording session, a tungsten microelectrode (~1.0MΩ @ 1 kHz; Frederick Haër & Co.) was lowered through a recording chamber and into the brain using a skull-mounted microdrive (MO-95, Narishige). Software, which was written in OpenEx (TDT Inc.), Labview (NI Inc.), and Matlab (The Mathworks Inc.), synchronized behavioral control with stimulus production and data collection. Neural signals were sampled at 24 kHz, amplified (RA16PA and

51

RZ2, TDT Inc.), and stored for online and offline analyses. Online spike sorting was conducted using OpenSorter (TDT Inc.).

3.3.5   Data-collection strategy.

While the electrode advanced through the brain, we presented white-noise bursts (duration: 100 ms; 65 dB SPL; 50 ms inter-tone-interval), which served as a 'search' stimulus to identify auditory-responsive sites. Once a responsive neuron was isolated, the monkey participated in the best-frequency task. A neuron was 'auditory' if the firing rate elicited by tone bursts was significantly (t-test, $p<0.05$) greater than the firing rate during a baseline silent period. 'Best frequency' was the frequency that elicited the largest response relative to a baseline period of silence. In those instances, when we could record multiple neurons from at a single site, the site's (and, hence, each neuron's) best frequency was calculated from the responses of the best-isolated single unit; typically, all of the neurons at a single recording site had comparable best frequencies. Only auditory neurons were further tested. Next, the monkey participated in the streaming task. On a trial-by-trial basis, tone B's frequency and listening time were randomly varied; the frequency of tone A was always set to the best frequency.

3.3.6   Behavioral analyses.

Behavioral analyses were similar to those that we reported earlier (Christison-Lagay et al., 2014a). We quantified the monkeys' performance by calculating the probability that the monkey reported hearing two streams, as a function of (1) the frequency difference between tones A and B and (2) listening time. The monkeys' performance for a particular frequency difference was considered significant when it did not overlap with chance performance (i.e., 0.5; Wilcoxon signed-rank test, $p<0.05$).

Because the monkeys' performance was variable, we developed a bootstrap procedure to establish the monkeys' performance thresholds and identify runs of trials that exceeded these thresholds (Christison-Lagay et al., 2014a). The bootstrap procedure generated a distribution of simulated data that shuffled the order of trials but maintained the relationship between a particular auditory stimulus and a monkey's response. This bootstrap procedure tested whether, within an experimental session, there were temporal epochs or 'runs' of trials that were above chance: for each experimental session, we identified those trials in which the frequency difference between tones A and B was 0.5 or 12 semitones; we included these frequency differences because they generate consistent reports in human listeners (Bregman, 1990; Cusack, 2005; Elhilali et al., 2009; Micheyl et al., 2007). We did not include 3 and 5 semitone frequency differences because they do not generate consistent reports in human listeners: there is no 'incorrect' answer. While maintaining the relationship between the stimuli and response, we shuffled the order of the trials. This procedure maintained the relationship between the stimulus and response but disrupted the temporal order of these stimulus-response pairings. Then, to simulate the temporal dynamics of a behavioral session, we analyzed performance as a function 20 consecutive shuffled stimulus-response pairings. This procedure was repeated 1000 times for each behavioral session. From this procedure, we generated a distribution of running averages. Finally, we calculated the 'running-average window (RAW) threshold'. We calculated the RAW threshold from each session's running-average distribution: the RAW threshold was defined as the upper boundary of the distribution's 95% confidence interval.

To compare the monkeys' performance with the bootstrapped performance, we extracted consecutive blocks of data that contained 20 trials in which the frequency difference was 0.5 or 12 semitones. Because the actual dataset contained trials from all of the tested-frequency

differences, the actual length of the data block could be longer than the window size. For example, for the window size of 20 trials, a data block might contain 23 trials: 20 trials in which the frequency difference was 0.5 or 12 semitones and 3 trials in which the frequency difference was 3 or 5 semitones. When the monkey's performance on the 0.5 and 12 semitone trials exceeded the RAW threshold, the entire trial block (including trials in which the frequency difference was 3 or 5 semitones) was considered 'suprathreshold'; the determination of 'suprathreshold' was only based on the performance during the 0.5- and 12-semitone trials because only these trial types were used in the bootstrap procedure. Using the suprathreshold data, we calculated, as a function of each frequency difference, the probability that the monkey reported two streams.

3.3.7   Neural analyses.

 Neural signals were re-sorted offline into individual single units using an automatic spike-sorting procedure (Quian Quiroga et al., 2006; Tsunada et al., 2011a). Data are reported in terms of average firing rate per tone burst. Data were aligned relative the onset of each auditory-stimulus sequence and each neuron's response latency. Additionally, because each stimulus sequence had a different (listening) duration, analyses were restricted to the time period encompassed by the first 12 tone bursts, which captured 68% of the data across all of the recording sessions.

Neural analyses to test relationship between neural activity and choice behavior.

Choice probability quantifies the ability of an ROC-based ideal observer to use spiking activity to discriminate choices for identical stimuli (Britten et al., 1996; Gu et al., 2007; Purushothaman et al., 2005; Russ et al., 2008a; Tsunada et al., 2011a). On a neuron-by-neuron and a tone-burst-by-tone-burst basis and as a function of semitone separation, we conducted this

ROC analysis after forming two distributions of firing-rate values based on a monkey's reports (i.e., '1 auditory stream' versus '2 auditory streams'). A choice-probability value of 0.5 indicates that an ideal observer could not use firing rate to distinguish between reports of 'one stream' and 'two streams'; whereas a choice-probability value of 1.0 indicates that an ideal observer could perfectly predict, using firing rate alone, whether a monkey reported 'one stream' or 'two streams'.

3.4   RESULTS

3.4.1   Psychophysical performance

Monkeys H and S reported whether a temporal sequence of tone bursts, in which the tone bursts alternated between two frequencies (see Fig. 3-1B), was heard as one or two auditory streams. The results from 108 sessions (Monkey H: 61 sessions; Monkey S: 47 sessions) are shown in Fig. 3-2. Because the monkeys had comparable performance, we pooled their behavioral data. These data were only from the recording sessions reported here and reproduce our previous behavioral findings (Christison-Lagay et al., 2014a). Fig. 3-2 plots the probability that the monkeys reported 'two auditory streams' as a function of the semitone separation between tones A and B. When the frequency difference was 0.5 semitones, the monkeys' performance was significantly (probability of two-stream reports=0.368±0.02, Wilcoxon signed-rank test, $p<0.05$) less than chance, indicating they were more likely to report '1 auditory stream'. In contrast, when the semitone separation was 12 semitones, the monkeys reliably (probability of two-stream reports=0.577±0.021, Wilcoxon signed-rank test, $p<0.05$) reported '2 auditory streams'. When the frequency difference was 3 or 5 semitones, the monkeys' performance was intermediate between these two extreme semitone values. That is, although they

55

reliably (3 semitone-tone separation: probability of two-stream reports=0.419±0.026, Wilcoxon signed-rank test, $p<0.05$; 5 semitone-tone separation: probability of two-stream reports=0.398±0.023, Wilcoxon signed-rank test, $p<0.05$) reported '1 auditory stream', they were significantly (Wilcoxon rank-sum test, $p<0.05$) more likely to report '2 auditory streams' relative to the 0.5-semitone condition and significantly (Wilcoxon rank-sum test, $p<0.05$) less likely to report '2 auditory streams' relative to the 12-semitones condition.

Although it is clear that the monkeys' struggled with this difficult task, we observed short runs of trials in which the monkeys clearly were performing well. To quantify this observation, we analyzed the monkeys' behavior relative to a bootstrapped simulation. In this simulation, we maintained the relationship between a stimulus-response pairing (e.g., 0.5 semitones and a report of '1 auditory stream') but shuffled the temporal order of these pairings. This procedure quantified explicitly whether there were significant short runs of performance that exceeded chance. We found that the monkeys' performance, relative to this bootstrapped distribution (i.e., the RAW threshold), increased modestly for 0.5- and 12-semitone trials (Fig. 3-2B). For intermediate semitone separations, behavioral performance across times of high performance, relative to overall performance, was the same, indicating that behavioral performance was stable across each session. Overall, the monkeys' pattern of behavior was consistent with—albeit poorer than—human performance (Christison-Lagay et al., 2014a). We emphasize that this is the first time that a non-human animal has been trained to directly report their auditory-streaming percepts in a manner comparable to human listeners (but see (Itatani et al., 2014; Izumi, 2002; Noda et al., 2013)).

3.4.2 Recording-site localization

Because A1 is the earliest stage of processing in the ventral auditory pathway, we focused on understanding how its neural responses might contribute to auditory perception. We isolated 108 A1 single units (Fig. 3-3A; 61 from monkey H and 47 from monkey S). Similar to previous work (Fu et al., 2004; Kajikawa et al., 2005; Kikuchi et al., 2010; O'Connell et al., 2014; Recanzone et al., 2000), A1 neurons were sharply frequency tuned (Fig. 3-3B and 3C) and had relatively short latencies (Fig. 3-3D). In our population, best frequencies ranged from 400 Hz to 3940 Hz; the median best frequency was 1984 Hz. Median latency (i.e., the first of two or more consecutive time bins that were >2 s.d. above a baseline period of silence) was 15 ms.

3.4.3 A1 responsivity during the streaming task: frequency sensitivity and habituation

During the streaming task, A1 neurons had auditory-driven spiking activity that was modulated by the frequency content and time course of the auditory sequence (single neuron: Fig. 3-4A; population activity: Fig. 3-4B). As expected, A1 neurons responded better to tone A, which was at a neuron's best frequency, than to the tone Bs, which were 0.5-12 semitones above this best frequency. Additionally, like previous reports (Fishman et al., 2004; Fishman et al., 2001a; Micheyl et al., 2005), the firing rate of A1 neurons habituated as the auditory sequence unfolded over time. However, as discussed below, unlike these previous reports, habituation was not frequency dependent.

Frequency tuning is broader during the streaming task than during the best-frequency task.

Fig. 3-5 plots the average frequency-response profiles during the streaming and best-frequency tasks. For the streaming-task data, we show the average firing rate in response to the first presentation of tone A (A_1) and tone B (B_1) from each semitone separation; we chose this

strategy to ensure that neural responses due to frequency tuning did not get conflated with changes in firing rate due to habituation (see below). For the best-frequency-task data, we only plot those frequency values that overlap with those presented during the streaming task. During the streaming task, we found that the average firing rate in response to the best frequency (i.e., 0 semitones) was significantly (2-factor ANOVA [task × frequency] and post-hoc tests, $p<0.05$) higher than the firing rates in response to the other frequency values and that the firing rates in response to these other frequency values were not significantly ($p>0.05$) different from one another. In contrast, during the best-frequency task, A1 firing rates generally decreased significantly ($p<0.05$) as frequency increased. Further, except at the best frequency, firing rates were significantly ($p<0.05$) higher during the streaming task than during the best-frequency task (see Supplemental Fig 3-1 for neuron-by-neuron comparison of normalized firing rates for the best frequency and streaming tasks). Together, these analyses indicate that A1 firing rates were less frequency selective during the streaming task than during the best-frequency task.

The firing rate of A1 neurons habituate—independent of frequency—as the auditory sequence unfolds over time.

Because, in previous reports, frequency-dependent habituation of A1 firing rate was proposed to be a coding mechanism for the streaming percept (Fishman et al., 2004; Fishman et al., 2001a; Micheyl et al., 2005), it was important to determine whether habituation manifested itself when our monkeys were actively reporting their streaming percepts. Indeed, we found that A1 firing rates habituated (single neuron: Fig. 3-4A; population activity: Fig. 3-4B). A1 neurons responded most vigorously to the first presentation of tone A (A_1). They were less responsive to the next tone burst (the first presentation of a tone B [B1]) and further decreases in firing rate with subsequent tone-burst presentations (A2…B6). Some individual A1 neurons displayed a

58

small amount of frequency-dependent habituation (e.g., see responses to tone B_1 in Fig. 3-4A). However, on average, this habituation was frequency independent (Fig. 3-4B), even for our most sharply tuned neurons (Fig. 3-4C). (A neuron was defined as 'sharply tuned' if it had a bandwidth of <1000 Hz at 25% of maximum firing rate; this was quantified by finding the two most disparate frequency bins in the neuron's tuning curve that elicited firing rates of >25% of the firing rate at the neuron's best frequency.)

This frequency-independent habituation can be seen most clearly in Fig. 3-4D, where we removed habituation's mean effect from each A1 response profile. To do this, we calculated the mean A1 response across all semitones (black line in Fig. 3-4B) and subtracted this from each neuron's response as a function of semitone separation. This subtraction procedure demonstrated that A1 spiking activity was not significantly (2-factor ANOVA [semitone difference × tone-burst position]; both main and interaction effects: $p > 0.05$) modulated by either semitone separation or 'tone-burst position' (i.e., tone burst A_1, tone burst B_1, etc.). This finding is consistent with the idea that A1 neurons during the streaming task had frequency-independent habituation.

To further quantify these observations, we used an ROC analysis (Green et al., 1966) to test the effects of semitone separation and habituation on A1 spiking activity. On a neuron-by-neuron basis, we calculated the ROC value between the average firing rate elicited by the first presentation of tone A and each subsequent tone burst ($B_1 A_2 \ldots B_6$). An ROC value of 0.5 indicates that an ideal observer could not distinguish between the firing rate elicited by tone A_1 and the firing rate elicited by any other tone burst in the sequence; whereas a value of 1 indicates that this observer could perfectly distinguish between these two responses. This analysis generally indicated that, independent of semitone separation, ROC values significantly (3-factor ANOVA with post-hoc tests [semitone difference × tone-burst position × neural population {all

versus sharply tuned}]; semitone-separation main effect: $p>0.05$; tone-burst position main effect: $p<0.05$; neural-population main effect: $p<0.05$; interactions: $p>0.05$) increased with tone-burst position and that sharply tuned neurons had larger ROC values than the overall population. Once again, this ROC analysis is consistent with a finding of frequency-independent habituation.

3.4.4   A1 neurons are modulated by the monkeys' choices

Finally, we tested choice probability (Fig. 3-7), which quantifies the ability of an ROC-based ideal observer to use A1 firing rate to discriminate between the monkeys' choices ('1 stream' versus '2 streams') for identical stimulus conditions. In this analysis, choice probability was calculated using the tone-burst-by-tone burst firing rates (i.e., not normalized to tone A_1's response). A three-factor ANOVA with post-hoc tests (frequency difference × tone-burst position × neural population [all versus sharply tuned versus supra-RAW threshold trials]; semitone-separation main effect: $p<0.05$; tone-burst-position main effect: $p<0.05$; neural-population main effect: $p<0.05$; frequency difference × tone-burst position: $p>0.05$; other interactions: $p<0.05$) indicated that, in general, as the auditory sequence unfolded over time, choice-probability values increased and became significantly different than chance following (on average) the onset of the fourth tone burst, peaking at ~0.75 and that choice probability was modulated by semitone difference. However, this latter effect was minimal because it was driven by the 5-semitone choice probability values, which tended to be slightly larger than the other values. Finally, sharply tuned A1 neurons had significantly ($p<0.05$) larger choice-probability values than the other two populations; whereas the choice-probability values from the supra-RAW-trials did not differ significantly ($p>0.05$) from the entire population, indicating that the monkeys' behavioral performance did not influence choice-probability values. Because the monkeys' choices can be

read out directly from their tone-burst-by-tone-burst firing rates, these results are contrary to previous predictions (Fishman et al., 2004; Fishman et al., 2001a; Micheyl et al., 2005) which suggested that frequency-dependent habituation was what encoded the monkeys' choices can be read out directly from, instead of necessarily requiring frequency-dependent habituation (for further data regarding choice-modulated activity, see Supplemental Fig. 3-2 for an example distribution of choice probability values and Supplemental Fig. 3-3 for the affect of choice on firing rate).

3.5   DISCUSSION

The principles underlying human listeners' perceptual organization of their acoustic environment have been elucidated by testing how they group and segregate auditory stimuli into one or more auditory streams (Bizley et al., 2013a; Bregman, 1990; Griffiths et al., 2004; Winkler et al., 2009). We recently reported that rhesus macaques stream auditory stimuli in a manner comparable to human listeners, with a task that uses the same criteria as human studies (Christison-Lagay et al., 2014a). Here, for the first time, we recorded A1 spiking activity while rhesus macaques reported their streaming percepts. Our most important finding was that A1 activity was modulated by the monkeys' reports of "1 auditory stream" versus "2 auditory streams"; this modulation increased with listening time and could be read out from the neurons' tone-by-tone firing rates. These findings contribute to our knowledge of how incoming auditory information is converted into a perceptual choice. However, because we find that tone-by-tone firing rate was modulated by choice, the findings are at odds with a prominent hypothesis by Micheyl et al. (2005), who argued that the streaming percept was encoded using a mechanism that required frequency-dependent habituation.

To understand why our findings are at odds with Micheyl et al. (2005), it is important to

61

first consider our finding that A1 frequency tuning was broader during our streaming task than during the best-frequency task (Fig. 3-5). This task-dependent tuning is consistent with previous work, demonstrating that A1 neurons sculpt their frequency sensitivity to the ongoing demands of a behavioral task (Aizenberg et al., 2013; Fritz et al., 2003; Fritz et al., 2012; Recanzone et al., 1993; Scheich et al., 2012; Shepard et al., 2012). In our case, because a neuron's best frequency (tone A) and a relatively wide range of frequencies above this value (the tone Bs) were all task relevant, we posit that A1 responsivity increased for all of these frequency values, resulting in broader frequency tuning relative to the best-frequency task. In other situations, if a task requires listeners to make fine frequency judgments, it can result in sharper frequency tuning (Recanzone et al., 1993).

Thus, one possible explanation for the difference between our finding and that of Micheyl et al. (2005) may relate to our aforementioned discussion of task-dependent frequency tuning. Specifically, we hypothesize that because in their study, the monkeys did not have to report their streaming percepts, the frequency tuning of their A1 neurons was relatively sharp and led to findings of frequency-specific rates of habituation. However, when the monkeys have to attend to the auditory sequence and report their streaming percepts, frequency tuning broadens, frequency-dependent habituation is eliminated (Figs. 3-4 through 3-6), and tone-burst-by-tone-burst firing rate emerges as a likely coding mechanism for the streaming percept. This broader frequency tuning during the active reporting of streaming percepts discounts a simple place code mechanism but is consistent with a mechanism that would use a combination of place and temporal code (Elhilali et al., 2009).

Broader frequency tuning may also explain the differences in tone masking observed between the current study and Fishman et al. (2004) and Fishman et al. (2001a). Fishman et al.

62

(2004) and Fishman et al. (2001a) reported that as tone B's frequency moved further from tone A's frequency, the more robust tone A's subsequent responses were; and inversely, that when tones A and B had similar frequencies, both tones elicited only small responses; this effect was attributed to forward masking (the exact mechanisms of which are not yet well-understood, but likely arise from a combination of post-synaptic inhibition and another mechanism such as synaptic depression (Wehr et al., 2005). Interestingly, although this forward-masked response is robust in Fishman et al, Micheyl et al. (2005) shows a much weaker modulation of tone A by tone B; and the current study finds no such modulation, but rather, reports that each successive tone presentation is smaller, and that frequency separation does not affect this response reduction. This is not necessarily to say that forward masking does not occur in the current study—indeed, masking has been shown to effect up to 5 seconds (Werner-Reiss et al., 2006): thus, masking could potentially affect responsivity throughout the duration of an entire trial. Furthermore, the specific effect of attention on forward-masking is not well understood: most studies have been preformed on anesthetized animals (Brosch et al., 1997; Brosch et al., 1999; Calford et al., 1995; Lu et al., 2000; Reale et al., 2000; Ulanovsky et al., 2003), and the few that have been performed on awake animals have shown variable affects of attention (Gottlieb et al., 1989; Werner-Reiss et al., 2006). Therefore, it is possible that our broader frequency tuning may also have affected the degree of frequency-dependence forward masking.

An additional caveat to the apparent discrepancy between our work and that of Micheyl et al. (2005) is that our auditory sequence and listening times were different than those reported by Micheyl et al. However, we do not believe that these differences can wholly account for the observed discrepancy. Indeed, because our analysis period already overlaps with the period of maximum neural habituation seen in Micheyl et al., it is unlikely that we would see the

emergence of frequency-dependent habituation even if our task had incorporated Micheyl et al.'s longer listening times. Nevertheless, it is still possible that if our study's task parameters more closely matched theirs, we might have seen frequency-dependent habituation. More work is needed to resolve this question and to determine whether an area downstream of A1 might use a coding mechanism like that suggested by Micheyl et al. (2005).

Another caveat to our findings is that while we report the mean choice probability values, the choice probability values are not always normally distributed (see Supplemental Fig. 3-2 for an example distribution of choice probability values). As can be seen, over time, an increasing portion of the population reaches a choice probability value of 1. Such values are generally not observed in either sensory or motor cortices, with typical choice probability values ranging from ~0.2 or 0.3 to ~0.7 or 0.8 (Bizley et al., 2013b; Cohen et al., 2009b; Heuer et al., 2004; Merten et al., 2013; Nienborg et al., 2014; Yang et al., 2010). Although having many neurons with choice probability values of 1 is unusual, several of the above mentioned studies, as well as another recent study in the primary auditory cortex, have neurons with choice probability values of 1 (Niwa et al., 2012b). It is not entirely clear why such a large subset of neurons in the current study yield high choice probability values, several possible explanations can be ruled out. We can rule out movement artifact as a possible explanation, as all of the data reported occurs before the monkeys were allowed to move the joystick (and trials with early movement were treated as errors and not used in the choice-probability analysis). Moreover, the effect cannot be explained as a regular loss of single-unit isolation over the course of a trial. First, trials of all types were pseudorandomly interleaved and behavior across the session was comparable, so loss of single-unit isolation should affect all choice probability values similarly. Furthermore, firing rates rebounded on a trial-by-trial basis (see Supplemental Fig. 3-4). The high choice probability

values cannot be attributed to faulty electrodes or other recording problems, as simultaneously recorded neurons at the same recording site show different choice probability values (see Supplemental Fig. 3-5). It is likely that the high values are due to habituation and a very short bin (40 ms). Under these conditions, the later tone presentations are increasingly likely to have very few, if any, spikes present for one of the choice conditions (Supplemental Fig. 3-3 shows how a behavioral report of 'two streams' is associated with lower firing rates).

The specific contribution of the auditory cortex to choice behavior is not clear. Consistent with our current findings (Figs. 3-6 and 3-7), recent studies (Bizley et al., 2013b; Niwa et al., 2013) have demonstrated that A1 has choice activity. In contrast, previous work from our group (Tsunada et al., 2011a; Tsunada et al., 2013) and others (Lemus et al., 2009) indicates that auditory-cortex neurons are not reliably modulated by choice. We suggest that one potential explanation for this apparent contradiction may be that tasks requiring decisions about relatively low-level stimulus features (e.g., pitch, amplitude modulation, stream segregation (Bizley et al., 2013b; Niwa et al., 2013), current report) might be represented directly (Nienborg et al., 2014) in the responses of individual neurons in the early ventral auditory pathway. In contrast, tasks that require a relatively high-level decision about the acoustic content of a stimulus (Lemus et al., 2009; Tsunada et al., 2011a; Tsunada et al., 2013) may be represented later in the ventral pathway.

One interpretation of A1 choice activity, like findings from other systems (Britten et al., 1996; Celebrini et al., 1994; Gu et al., 2007; Law et al., 2008), is that it reflects a feed-forward mechanism that uses A1 activity as sensory evidence for the eventual decision (Shadlen et al., 1996). Alternatively, these signals might represent feedback once the decision is formed elsewhere in ventral pathway (Nienborg et al., 2009; Niwa et al., 2013; Russ et al., 2008a). Future

work should focus on identifying the temporal window in which to conduct analyses that relate

neural activity with behavior in order to differentiate between these two possibilities (Cohen et

al., 2009a; Nienborg et al., 2009).
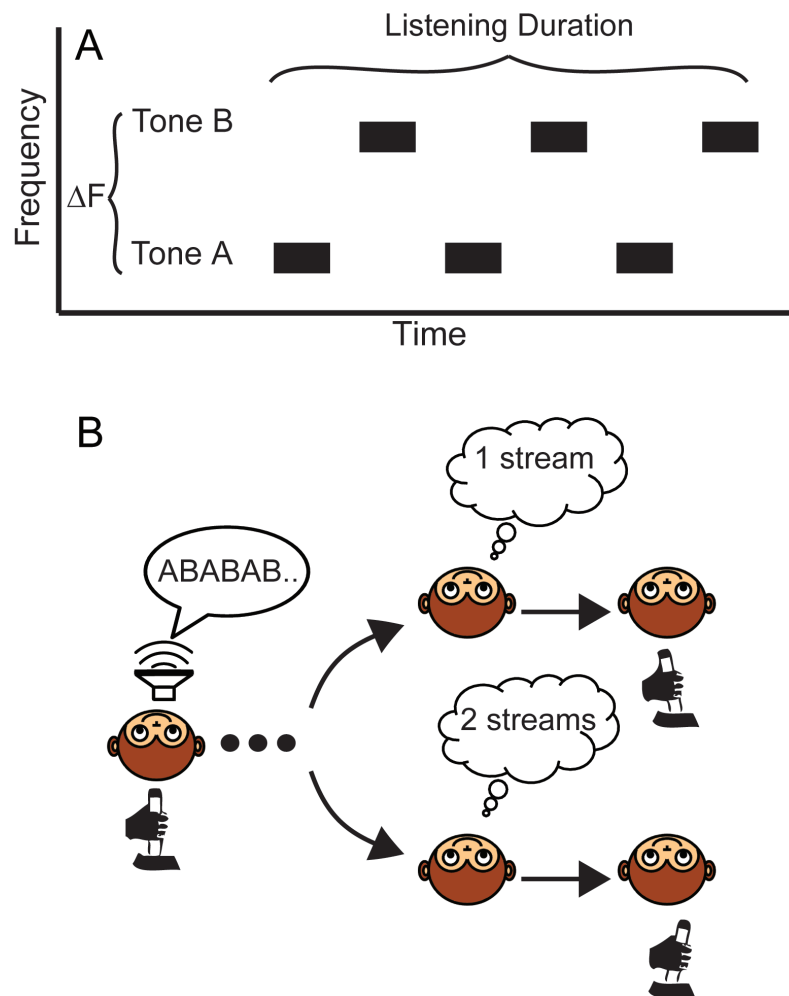
Figure 3-1: Task and stimulus.
(**A**) The monkey indicated his choice by moving a joystick to the right to report 'one auditory stream' or to the left to report 'two auditory streams'. The monkey made his report following offset of the auditory stimulus. (**B**) The auditory stimulus was a temporal sequence of tone bursts. On a trial-by-trial basis, we varied the frequency difference (ΔF) between tone A and tone B and the duration of the tone sequence (listening time); tone A was always at a neuron's best frequency.
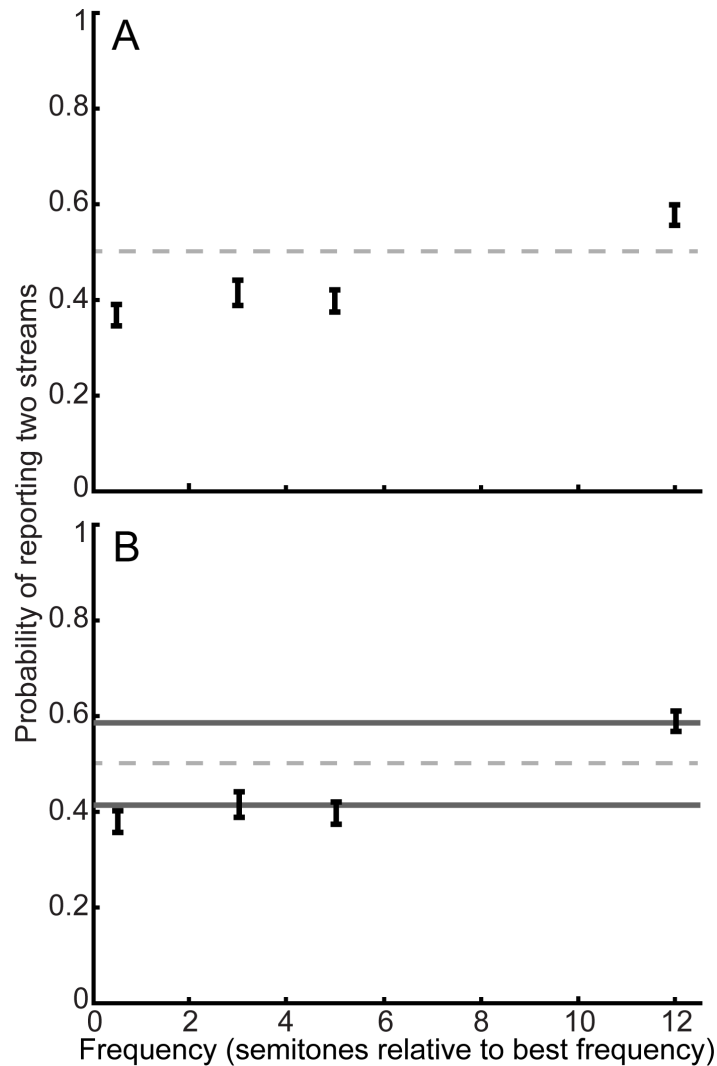
Figure 3-2: Psychophysical performance on the streaming task.
(**A**) Average psychometric performance for both monkeys is plotted as the proportion of trials in which the monkey reported "two streams" as a function of frequency (in semitones). 0 semitones represents each neuron's best frequency. The center of each bar indicates average performance. The length of the bars indicates the 95% confidence interval. The gray dashed line represents chance performance (0.5) of reporting one or two streams. (**B**) Psychometric behavior related to bootstrapped RAW threshold. This bootstrap procedue maintained the integrity between the stimulus and the monkeys' responses but shuffled the temporal order and tested explicitly whether there were significant temporal runs of high performance. RAW thresholds were calculated from session-by-session data. The center of each bar indicates average suprathreshold performance. The length of the bars indicates the 95% confidence interval. The solid line illustrates the upper and lower boundaries of the RAW threshold. The gray dashed line in represents chance performance (0.5) of answering one or two streams.
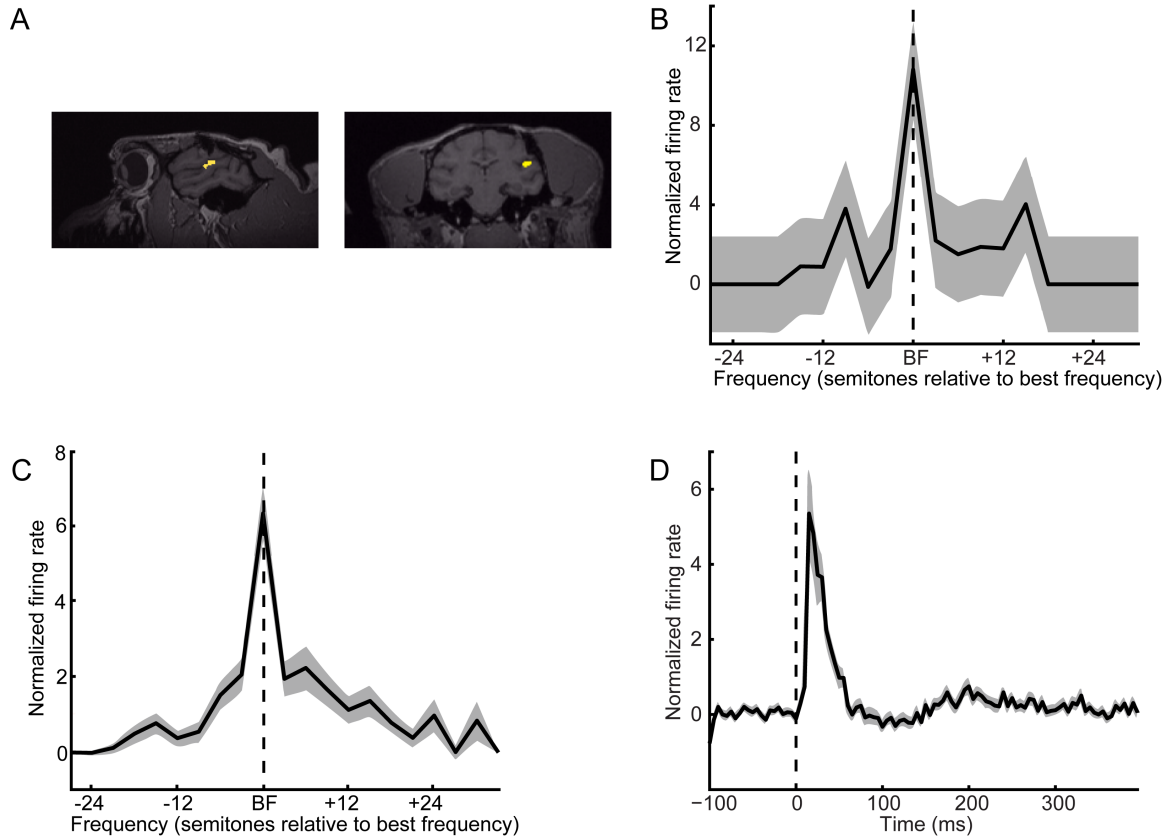
Figure 3-3: Recording locations and A1 response properties.
(**A**) Sagittal and coronal MRI sections of monkey H's brain at the level of the superior temporal gyrus. The yellow regions indicate the targeted location of A1. (**B**) Single-neuron and (**C**) population frequency-response profiles. These response profiles are plotted relative to a neuron's best frequency (BF). Vertical dotted lines indicate BF. (**D**) Population response profile. The vertical dotted line indicates stimulus onset of each tone burst. For all of the panels, firing rate is normalized relative to a baseline period of silence. Thick lines indicate mean values; shading indicates s.e.m.
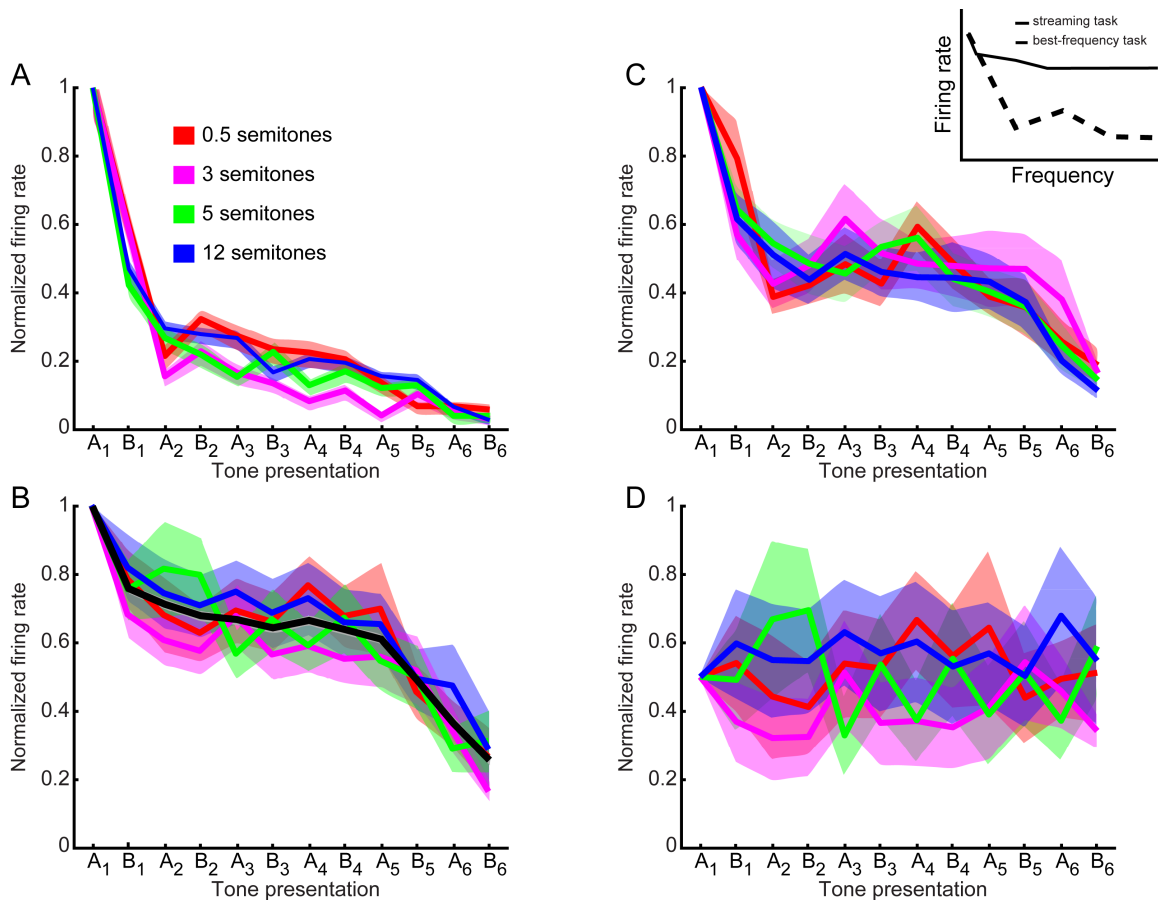
Figure 3-4: A1 sensitivity to stimulus frequency and tone presentation.
(**A**) Single-neuron example of A1 firing rate in response to the acoustic sequence; data are combined from reports of 'one stream' and 'two streams'. Color corresponds to semitone separation; see legend. Data are aligned relative to each tone burst in the sequence. The first tone burst is designated as 'A$_1$'; the second as 'B$_1$', the third as 'A$_2$' etc. (**B**) Population response profile, plotted as in **A**. The thick black line indicates the average response across all semitone separations. (**C**) Population response profile showing only neurons that are sharply tuned for frequency (i.e., those with a bandwidth of <1000 Hz at 25% of maximum firing rate; see main text for more details), plotted as in **A**. For panels **A**-**C**, firing rate is normalized relative to the mean firing rate elicited by tone A$_1$. (**D**) Population response profile in which the mean A1 response across all semitones (thick black line in **B**) was subtracted from each neuron's response as a function of semitone separation. Thick lines indicate mean values; shading indicates s.e.m. Inset at upper right is a schematic of the neural tuning during the streaming and best frequency tasks for reference; see Fig. 3-5 for full version.

70
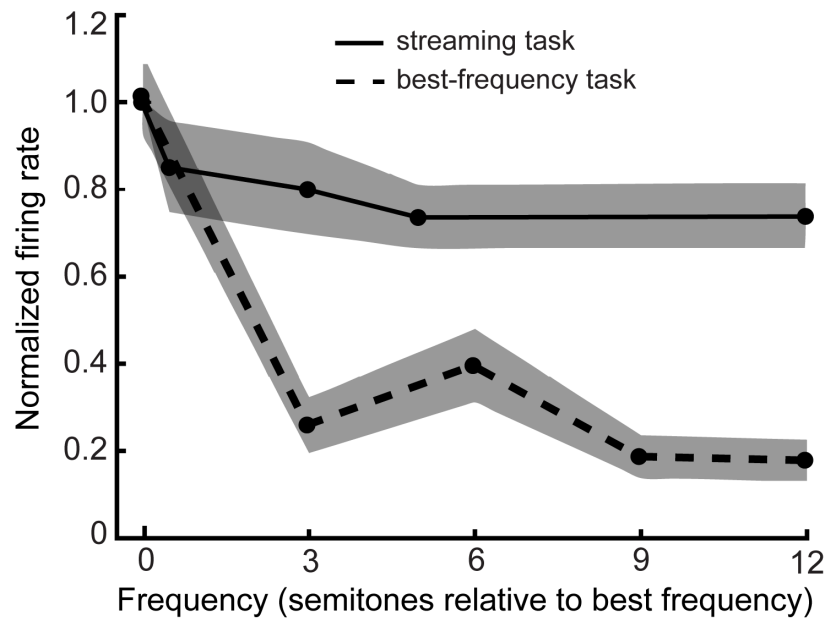
Figure 3-5: A1 frequency selectivity during streaming and best-frequency tasks.
Population frequency-response profiles during the streaming (solid line) and best-frequency (dashed line) tasks. The response profiles are restricted to frequency values common to those in both tasks. Thick lines indicate mean values; shading indicates s.e.m. To compare the sharpness of tuning across the two conditions, firing rate in the streaming task is normalized relative to the mean firing rate elicited by tone A_1 and relative to the mean firing rate elicited by the best frequency in the best-frequency task.
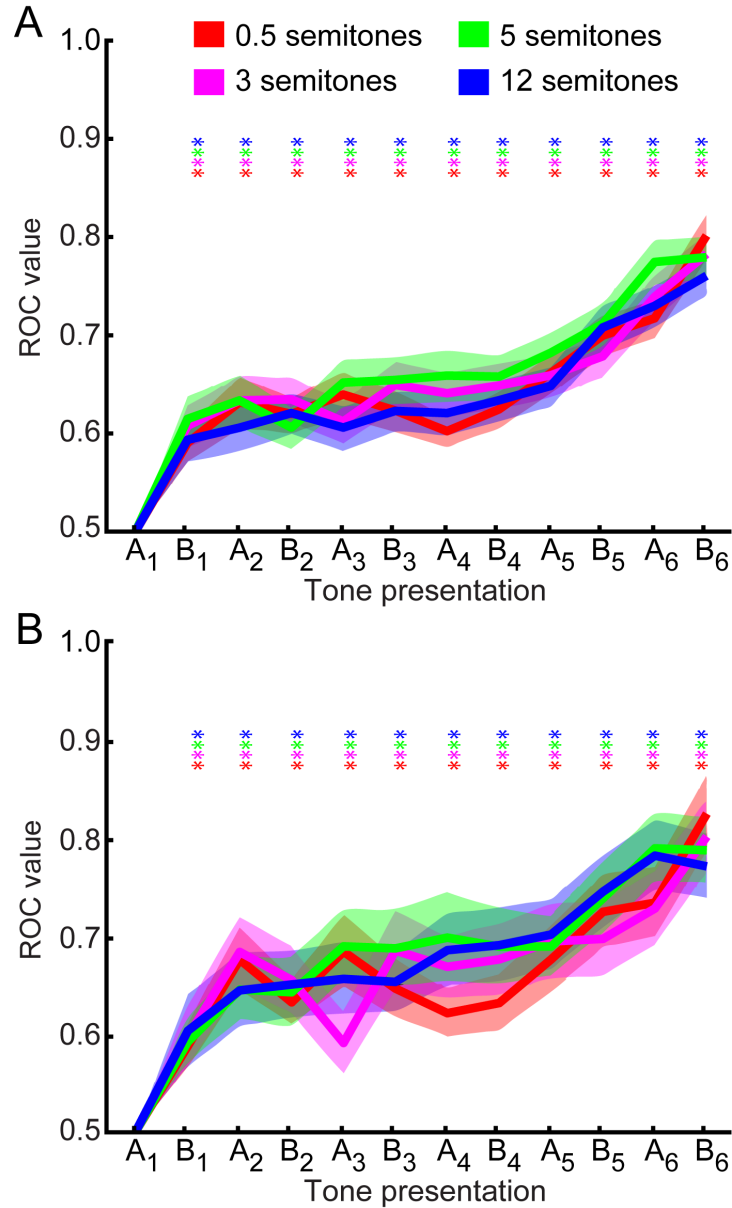
Figure 3-6: ROC analysis for A1 habituation sensitivity during streaming task.
ROC-based neural selectivity for tone bursts over time, relative to tone A_1's firing rate. Data are combined from reports of "one stream" and "two streams". The first tone burst is designated as "A_1"; the second as "B_1", the third as "A_2" etc. The entire population of neurons is shown in panel **A**, whereas panel **B** shows only neurons that are sharply tuned for frequency (i.e., those with a bandwidth of <1000 Hz at 25% of maximum firing rate; see main text for more details). Color corresponds to frequency difference; see legend. Thick lines indicate mean values; shading indicates s.e.m. Asterisks indicate mean ROC values that were significantly (0.5; t-test, $p < 0.05$; see legend) different than chance.
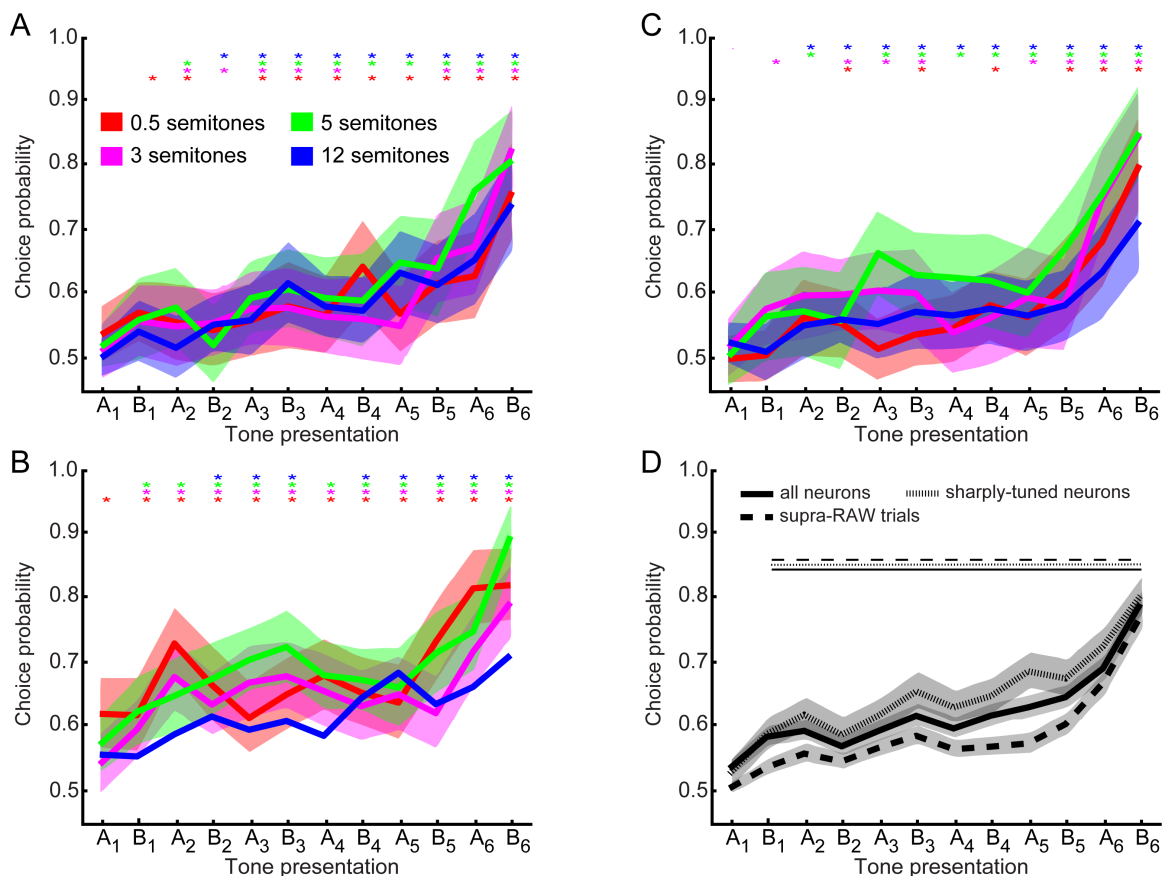
Figure 3-7: Choice selectivity of A1 neurons: tone-burst-by-tone-burst firing rate.
Distributions of choice probability relative to each tone burst in the sequence. Choice probability is calculated using mean firing rates elicited by each tone burst and not normalized as done in Fig. 3-7. The first tone burst is designated as 'A$_1$'; the second as 'B$_1$', the third as 'A$_2$' etc. Color corresponds to semitone separation; see legend. Panel **A** shows the entire population; panel **B** shows the subset of neurons that are sharply tuned for frequency (i.e., those with a bandwidth of <1000 Hz at 25% of maximum firing rate; see main text for more details); and panel **C** shows data from the subset of trials in which behavior exceeded the RAW thresholds. To more readily compare how choice probability evolved over time, we averaged the three populations of neurons' (**A**-**C**) choice-probability values across semitone separation (**D**). The entire population is plotted in the solid black line; sharply tuned neurons are shown in the hashed line; trials in which the RAW threshold was exceeded are shown in dashed line. Thick lines indicate mean values; shading indicates s.e.m. In panels **A**-**C**, asterisks indicate mean choice-probability values that were significantly different than chance (0.5; t-test, $p<0.05$; see legend). In panel **D**, horizontal lines indicate mean choice-probability values that were significantly different than chance (0.5; t-test, $p<0.05$; see legend).

73
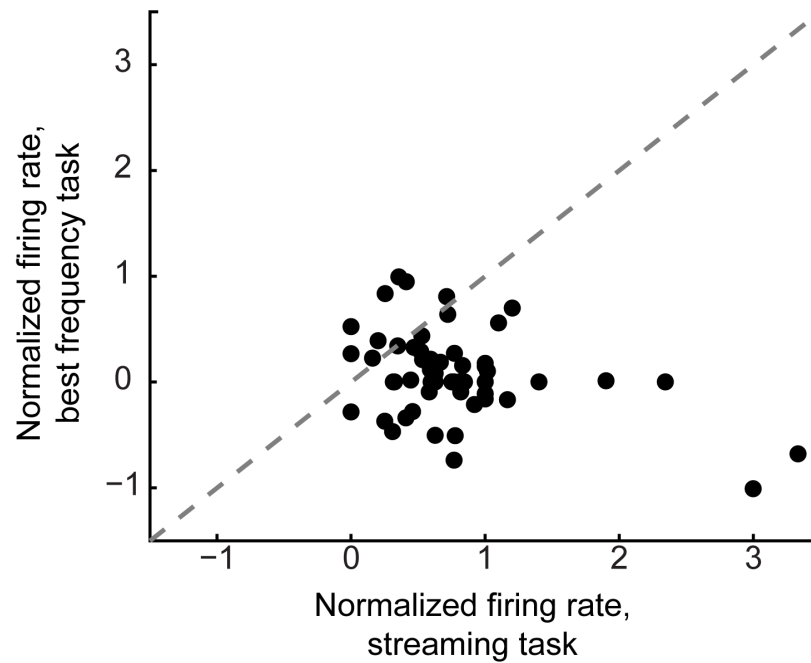
Supplemental Figure 3-1: Neuron-by-neuron frequency selectivity during streaming and best-frequency tasks.
Normalized firing rates for single neurons in response to tone B_1 in the streaming task (x-axis) are plotted against the normalized firing rate for single neurons to a 12 semitone tone burst in the best-frequency task. 12 semitones was chosen because it was used in both the streaming and best-frequency tasks. Gray dotted line shows the line of unity. The firing rates elicited during the best-frequency task were significantly lower than those elicited during the streaming task (t-test, $p<0.05$).
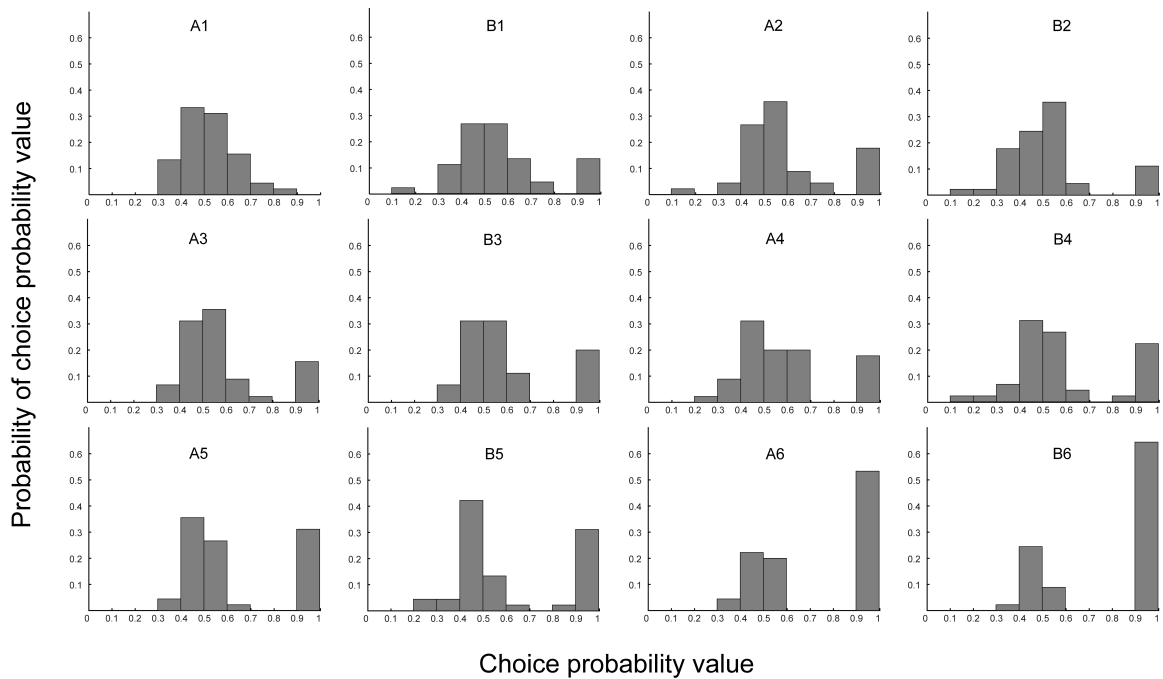
Supplemental Figure 3-2: Example distribution of choice probability values: 5 semitone trials

The distributions of choice probability relative to each tone burst in the sequence for 5 semitones. The first tone burst is designated as 'A$_1$'; the second as 'B$_1$', the third as 'A$_2$' etc. This distribution was used to calculate the population choice probability values, as shown in Fig. 3-7, panel A.
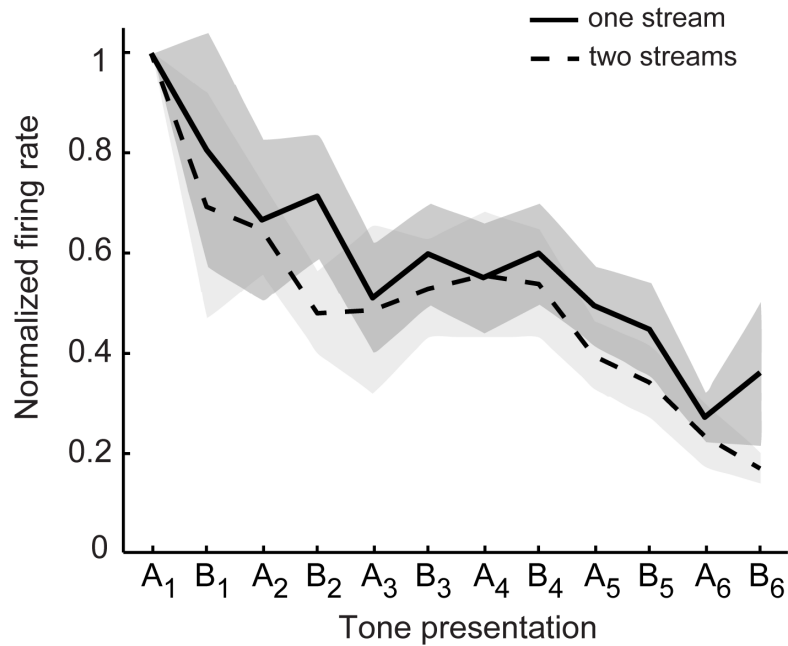
Supplemental Figure 3-3 Response profile as a function of choice during the streaming task.
Population response profile of A1 firing rate in response to the acoustic sequence for 5 semitone trials; data are separated by reports of 'one stream' and 'two streams'. The solid line indicates the average firing rate for choosing one stream; the dotted line indicates the average firing rate for choosing two streams; thick lines indicate mean values; shading indicates s.e.m. Data are aligned relative to each tone burst in the sequence. The first tone burst is designated as 'A$_1$'; the second as 'B$_1$', the third as 'A$_2$' etc.
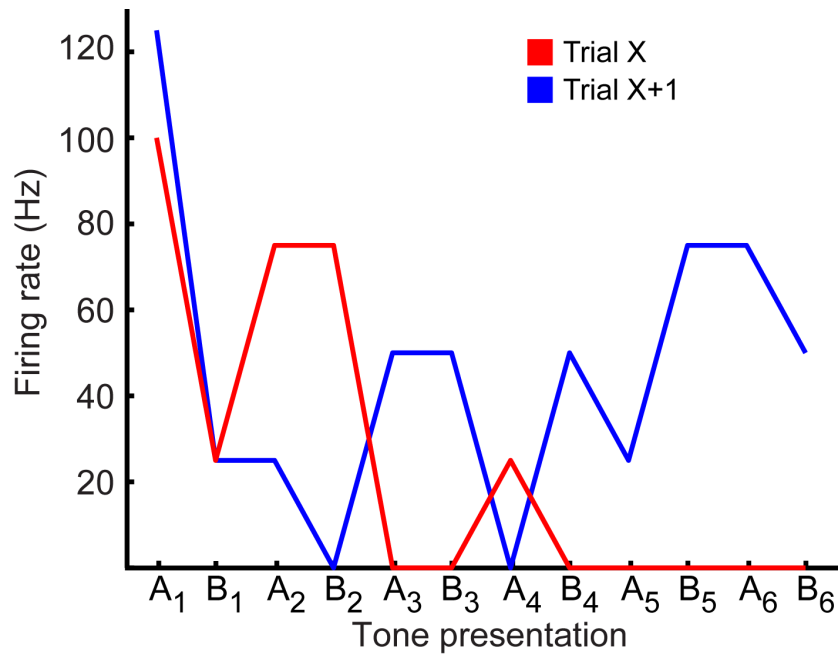
Supplemental Figure 3-4 Successive trials show recovery from habituation.
Representative example of a neuron's firing rate to successive trials of the same type. Colors
indicate trial number. Even though the neuron ceases to fire in response to the tones for the latter
tone bursts (shown in red), the blue line indicates that the neuron began firing again at the start of
the next trial. Data are aligned relative to each tone burst in the sequence. The first tone burst is
designated as 'A$_1$'; the second as 'B$_1$', the third as 'A$_2$' etc.
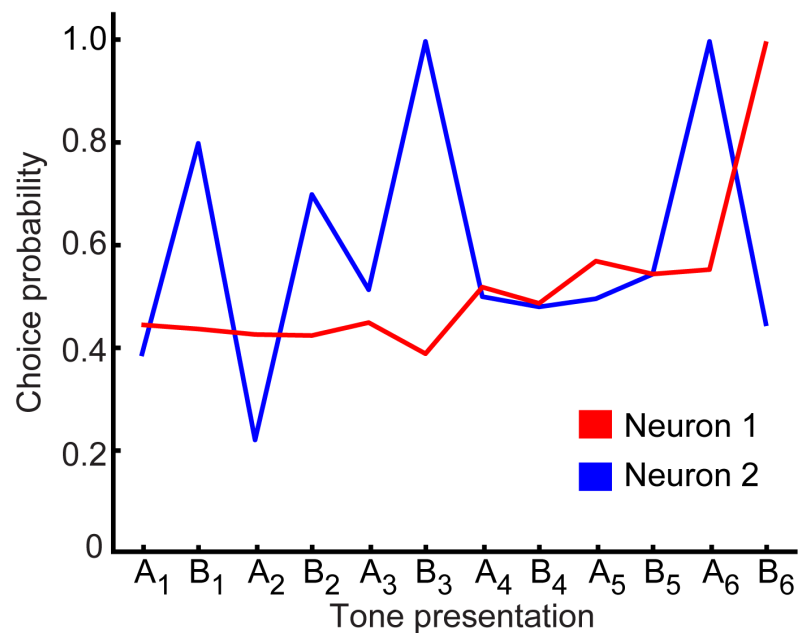
Supplemental Figure 3-5 Neurons simultaneously recorded from the same site show different choice modulation.
Example of the choice probabilities from two neurons recorded simultaneously from the same site shows that neurons at the same site may exhibit different choice-related activity. Shown in the response to 12-semitone trials. Colors indicate neuron identity. Data are aligned relative to each tone burst in the sequence. The first tone burst is designated as 'A$_1$'; the second as 'B$_1$', the third as 'A$_2$' etc.

3.8 REFERENCES

Aizenberg, M., Geffen, M.N. 2013. Bidirectional effects of aversive learning on perceptual acuity are mediated by the sensory cortex. Nat Neurosci 16, 994-6.

Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D. 2004. Neural correlates of sensory and decision processes in auditory object identification. Nature neuroscience 7, 295-301.

Bizley, J.K., Cohen, Y.E. 2013a. The what, where and how of auditory-object perception. Nat Rev Neurosci 14, 693-707.

Bizley, J.K., Walker, K.M., Nodal, F.R., King, A.J., Schnupp, J.W. 2013b. Auditory cortex represents both pitch judgments and the corresponding acoustic cues. Curr Biol 23, 620-5.

Bregman, A.S. 1990. Auditory Scene Analysis: The Perceptual Organization of Sound MIT Press, Cambridge, MA.

Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S., Movshon, J.A. 1996. A relationship between behavioral choice and the visual responses of neurons in macaque MT. Visual Neuroscience 13, 87-100.

Brosch, M., Schreiner, C.E. 1997. Time course of forward masking tuning curves in cat primary auditory cortex. J Neurophysiol 77, 923-43.

Brosch, M., Schulz, A., Scheich, H. 1999. Processing of sound sequences in macaque auditory cortex: response enhancement. J Neurophysiol 82, 1542-59.

Calford, M.B., Semple, M.N. 1995. Monaural inhibition in cat auditory cortex. J Neurophysiol 73, 1876-91.

Celebrini, S., Newsome, W.T. 1994. Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. Journal of Neuroscience 14, 4109-24.

Christison-Lagay, K.L., Cohen, Y.E. 2014a. Behavioral correlates of auditory streaming in rhesus macaques. Hear Res 309, 17-25.

Cohen, M.R., Newsome, W.T. 2009a. Estimates of the Contribution of Single Neurons to Perception Depend on Timescale and Noise Correlation. J Neurosci 29, 6635-6648.

Cohen, M.R., Maunsell, J.H. 2009b. Attention improves performance primarily by reducing interneuronal correlations. Nat Neurosci 12, 1594-600.

Cusack, R. 2005. The Intraparietal Sulcus and Perceptual Organization. J Cogn Neurosci 17, 641-651.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A.J., Shamma, S.A. 2009. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. Neuron 61, 317-29.

Fishman, Y.I., Arezzo, J.C., Steinschneider, M. 2004. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J Acoust Soc Am 116, 1656-70.

Fishman, Y.I., Reser, D.H., Arezzo, J.C., Steinschneider, M. 2001a. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear Res 151, 167-187.

Fritz, J., Shamma, S., Elhilali, M., Klein, D. 2003. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nature neuroscience 6, 1216-23.

Fritz, J.B., David, S.V., Shamma, S.A. 2012. Attention and dynamic, task-related receptive field plasticity in adult auditory cortex. In: Cohen, Y.E., Fay, R.R., Popper, A.N., (Eds.),

Neural Correlates of Auditory Cognition. Springer-Verlag. pp. 251-291

Fu, K.M., Shah, A.S., O'Connell, M.N., McGinnis, T., Eckholdt, H., Lakatos, P., Smiley, J., Schroeder, C.E. 2004. Timing and laminar profile of eye-position effects on auditory responses in primate auditory cortex. J Neurophysiol 92, 3522-31.

Giordano, B.L., McAdams, S., Kriegeskorte, N., Zatorre, R.J., Belin, P. 2012. Abstract encoding of auditory objects in cortical activity patterns. Cereb Cortex.

Gottlieb, Y., Vaadia, E., Abeles, M. 1989. Single unit activity in the auditory cortex of a monkey performing a short term memory task. Exp Brain Res 74, 139-48.

Green, D., Swets, J. 1966. Signal Detection Theory and Psychophysics John Wiley and Sons, Inc, New York.

Griffiths, T.D., Warren, J.D. 2004. What is an auditory object? Nat Rev Neurosci 5, 887-892.

Gu, Y., DeAngelis, G.C., Angelaki, D.E. 2007. A functional link between area MSTd and heading perception based on vestibular signals. Nat Neurosci 10, 1038-47.

Gutschalk, A., Micheyl, C., Melcher, J.R., Rupp, A., Scherg, M., Oxenham, A.J. 2005. Neuromagnetic correlates of streaming in human auditory cortex. J Neurosci 25, 5382-8.

Hackett, T.A. 2011. Information flow in the auditory cortical network. Hear Res 271, 133-46.

Heuer, H.W., Britten, K.H. 2004. Optic flow signals in extrastriate area MST: comparison of perceptual and neuronal sensitivity. J Neurophysiol 91, 1314-26.

Itatani, N., Klump, G.M. 2014. Neural correlates of auditory streaming in an objective behavioral task. Proc Natl Acad Sci U S A 111, 10738-43.

Izumi, A. 2002. Auditory stream segregation in Japanese monkeys. Cognition 82, B113-B122.

Kajikawa, Y., de La Mothe, L., Blumell, S., Hackett, T.A. 2005. A comparison of neuron response properties in areas A1 and CM of the marmoset monkey auditory cortex: tones and broadband noise. J Neurophysiol 93, 22-34.

Kajikawa, Y., Camalier, C.R., de la Mothe, L.A., D'Angelo, W.R., Sterbing-D'Angelo, S.J., Hackett, T.A. 2011. Auditory cortical tuning to band-pass noise in primate A1 and CM: A comparison to pure tones. Neurosci Res 70, 401-407.

Kikuchi, Y., Horwitz, B., Mishkin, M. 2010. Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. J Neurosci 30, 13021-30.

Kusmierek, P., Rauschecker, J.P. 2009. Functional specialization of medial auditory belt cortex in the alert rhesus monkey. J Neurophysiol 102, 1606-22.

Law, C.T., Gold, J.I. 2008. Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. Nat Neurosci 11, 505-13.

Lemus, L., Hernandez, A., Romo, R. 2009. Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. Proc Natl Acad Sci U S A 106, 9471-6.

Lu, T., Wang, X. 2000. Temporal discharge patterns evoked by rapid sequences of wide- and narrowband clicks in the primary auditory cortex of cat. J Neurophysiol 84, 236-46.

McAdams, S., Bregman, A.S. 1979. Hearing musical streams. Compt. Music J. 3, 26-63.

Merten, K., Nieder, A. 2013. Comparison of abstract decision encoding in the monkey prefrontal cortex, the presupplementary, and cingulate motor areas. J Neurophysiol 110, 19-32.

Mesgarani, N., Chang, E.F. 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485, 233-6.

Micheyl, C., Tian, B., Carlyon, R.P., Rauschecker, J.P. 2005. Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron 48, 139-48.

Micheyl, C., Carlyon, R.P., Gutschalk, A., Melcher, J.R., Oxenham, A.J., Rauschecker, J.P., Tian, B., Courtenay Wilson, E. 2007. The role of auditory cortex in the formation of auditory streams. Hear Res 229, 116-31.

Nienborg, H., Cumming, B.G. 2009. Decision-related activity in sensory neurons reflects more than a neuron's causal effect. Nature 459, 89-92.

Nienborg, H., Cumming, B.G. 2014. Decision-related activity in sensory neurons may depend on the columnar architecture of cerebral cortex. J Neurosci 34, 3579-85.

Niwa, M., Johnson, J.S., O'Connor, K.N., Sutter, M.L. 2012a. Active engagement improves primary auditory cortical neurons' ability to discriminate temporal modulation. J Neurosci 32, 9323-34.

Niwa, M., Johnson, J.S., O'Connor, K.N., Sutter, M.L. 2012b. Activity related to perceptual judgment and action in primary auditory cortex. J Neurosci 32, 3193-210.

Niwa, M., Johnson, J.S., O'Connor, K.N., Sutter, M.L. 2013. Differences between Primary Auditory Cortex and Auditory Belt Related to Encoding and Choice for AM Sounds. J Neurosci 33, 8378-95.

Noda, T., Kanzaki, R., Takahashi, H. 2013. Stimulus phase locking of cortical oscillation for auditory stream segregation in rats. PLoS One 8, e83544.

O'Connell, M.N., Barczak, A., Schroeder, C.E., Lakatos, P. 2014. Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. J Neurosci 34, 16496-508.

Purushothaman, G., Bradley, D.C. 2005. Neural population code for fine perceptual decisions in area MT. Nat Neurosci 8, 99-106.

Quian Quiroga, R., Snyder, L.H., Batista, A.P., Cui, H., Andersen, R.A. 2006. Movement intention is better predicted than attention in the posterior parietal cortex. J Neurosci 26, 3615-20.

Rauschecker, J.P. 2012. Ventral and dorsal streams in the evolution of speech and language. Frontiers in evolutionary neuroscience 4, 7.

Rauschecker, J.P., Tian, B. 2004. Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. J Neurophysiol 91, 2578-89.

Rauschecker, J.P., Scott, S.K. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12, 718-24.

Reale, R.A., Brugge, J.F. 2000. Directional sensitivity of neurons in the primary auditory (AI) cortex of the cat to successive sounds ordered in time and space. J Neurophysiol 84, 435-50.

Recanzone, G.H., Schreiner, C.E., Merzenich, M.M. 1993. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. J Neurosci 13, 87-103.

Recanzone, G.H., Guard, D.C., Phan, M.L. 2000. Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. J Neurophysiol 83, 2315-2331.

Romanski, L.M., Averbeck, B.B. 2009. The primate cortical auditory system and neural representation of conspecific vocalizations. Annu Rev Neurosci 32, 315-46.

Russ, B.E., Orr, L.E., Cohen, Y.E. 2008a. Prefrontal neurons predict choices during an auditory same-different task. Curr Biol 18, 1483-8.

Scheich, H., Brosch, M. 2012. Task-Related Activation of Auditory Cortex. In: Cohen, Y.E., Fay, R.R., Popper, A.N., (Eds.), Neural Correlates of Auditory Cognition. Springer-Verlag. pp. 45-81.

Shadlen, M.N., Britten, K.H., Newsome, W.T., Movshon, J.A. 1996. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. Journal of Neuroscience 16, 1486-510.

Shepard, K.N., Kilgard, M.P., Liu, R.C. 2012. Experience-dependent Plasticity and Auditory Cortex. In: Cohen, Y.E., Fay, R.R., Popper, A.N., (Eds.), Neural Correlates of Auditory Cognition. Springer-Verlag. pp. 293-327.

Tsunada, J., Lee, J.H., Cohen, Y.E. 2011a. Representation of speech categories in the primate auditory cortex. J Neurophysiol.

Tsunada, J., Liu, A.S., Gold, J.I., Cohen, Y.E. 2013. Causal role of primate auditory cortex in auditory perceptual decision-making, Annual Meeting of the Society for Neuroscience, San Diego, CA. pp. 636.01.

Ulanovsky, N., Las, L., Nelken, I. 2003. Processing of low-probability sounds by cortical neurons. Nat Neurosci 6, 391-398.

Wehr, M., Zador, A.M. 2005. Synaptic mechanisms of forward suppression in rat auditory cortex. Neuron 47, 437-45.

Werner-Reiss, U., Porter, K.K., Underhill, A.M., Groh, J.M. 2006. Long lasting attenuation by prior sounds in auditory cortex of awake primates. Exp Brain Res 168, 272-6.

Winkler, I., Denham, S.L., Nelken, I. 2009. Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci 13, 532-40.

Yang, S.N., Hwang, H., Ford, J., Heinen, S. 2010. Supplementary eye field activity reflects a decision rule governing smooth pursuit but not the decision. J Neurophysiol 103, 2458-69.

4   CHAPTER 4: DISCUSSION

In this dissertation, we have examined the neural and behavioral correlates of auditory streaming in rhesus macaques. Our results demonstrated for the first time that rhesus macaques' behavioral reports were qualitatively consistent with those of human listeners. Specifically, we found that their reports were modulated by frequency separation, listening duration, and temporal overlap in a manner consistent with humans and that their behavioral reports were independent of the absolute frequency content of the stimulus. We also found that, like previous studies (Fishman et al., 2004; Fishman et al., 2001a; Micheyl et al., 2005), A1 neurons had frequency-tuned responses that habituated as the auditory sequence unfolded over time. More significantly, we showed for the first time that firing rate of A1 neurons was modulated by the monkeys' choices. These findings provide the first direct evidence that (1) monkeys stream auditory stimuli in a manner consistent with human listeners, and (2) A1 activity can contribute to the sensory evidence underlying the segregation and grouping of acoustic stimuli into distinct auditory streams. In this chapter, we discuss the further implications of our findings, caveats to the current studies, challenges faced in the course of the studies, and future directions.

4.1   THE IMPORTANCE AND CHALLENGES OF USING THE STREAMING TASK

Results from previous studies suggest that humans and animals stream auditory stimuli in a comparable manner (Itatani et al., 2014; Izumi, 2002; Ma et al., 2010; Moerel et al., 2012; Noda et al., 2013; Russ et al., 2008a; Tsunada et al., 2011a), but these studies used tasks that indirectly measured streaming and assumed that animals stream sounds like human listeners. This assumption is not unreasonable because humans and animals have similar auditory perceptual abilities (Izumi, 2002; Kuhl et al., 1975a; Kuhl et al., 1982; Petkov et al., 2003; Petkov et al., 2007; Recanzone et al., 2008). However, until it is demonstrated that humans and animals stream

83

auditory stimuli in a comparable manner, the use of these indirect measures presents a potential problem: if animals and humans do *not* stream comparably, then these indirect measures would not be a valid model for human hearing. To avoid this potential fallacy, comparable methodologies must be used to explicitly test human versus animal auditory perception.

Although directly validating monkeys as a model human streaming was important, it was difficult to implement the standard methodologies used in human steaming studies. This was foremost because the behavioral task was incredibly challenging for the animals. This manifested in two ways: a 3+ year time to train the animals on the task and difficulty with motivating the monkeys to complete large numbers of trials. It is unclear exactly why the task was difficult for monkeys, as animals have been successfully trained on a number of complex tasks (including in our lab (Christison-Lagay et al., 2014b; Russ et al., 2007; Tsunada et al., 2011a)).

One possible explanation is that the streaming percept is somewhat abstract (even to human listeners). To illustrate this point: regardless of whether one or two streams is perceived, human listeners always report hearing two frequencies—it's simply that sometimes the two frequencies in the auditory sequence merge together into a single percept that sounds like 'galloping', and sometimes the listeners distinctly hear two streams that are composed of tone bursts with different frequencies (Bregman, 1990). Thus, a successfully trained monkey listener must learn the difficult concept of reporting the number of *streams*, not the number of frequencies; see Chapter 2 for control analyses that indicate that the monkeys were not simply reporting the number of frequencies or the relative frequency difference.

Despite the apparent difficulty of the task, we noted time periods when the monkeys performed quite well. To test this observation, we developed the RAW thresholds (see Chapters 2 and 3 for further description) to extract epochs of good performance. As discussed in Chapter 2,

monkeys' performance during these 'on task' trials (those that are supra-RAW threshold) clearly

indicates that monkeys reports are comparable to human reports. It is important to note that while

the supra-RAW threshold trials show this conclusion most clearly, their overall performance was

also consistent with human behavior.

4.2   THE NEURAL ENCODING OF AUDITORY STREAMS

Previous neurophysiology studies of auditory streaming have recorded neural activity in

A1 either during passive-listening tasks (Fishman et al., 2004; Fishman et al., 2001a) or during

active-listening conditions in tasks that were not directly related to auditory streaming (Lakatos et

al., 2013; Micheyl et al., 2005). These studies provide important insights into A1 activity, but

they could not offer direct insights into whether A1 activity codes the monkeys streaming

percepts.

Our study (Chapter 3) is the first study to directly test the relationship between A1

spiking activity and streaming percepts. Consistent with previous studies, we found that A1

neurons had frequency-tuned responses that habituated (Fishman et al., 2004; Fishman et al.,

2001a; Micheyl et al., 2005); and additionally found, for the *first time*, that activity in A1 neurons

was modulated by the monkeys' choices.

To put our finding of choice activity in A1 into context, Micheyl et al. (2005) proposed

that habituation in A1 neurons could encode perceptual choice, and Elhilali et al. (2009) proposed

that temporal coherence of activity across regions of A1 could be a correlate of perceptual choice.

Although the specifics of our results are at odds with Micheyl et al. (2005)'s hypothesis (i.e., we

find that tone-by-tone firing rate alone, rather than frequency-dependent habituation, is sufficient

to encode choice, see Chapter 3 for further discussion), our data provide the first direct evidence

that A1 contributes to encoding the streaming percept. Similarly, several recent studies have

85

found that A1 activity is modulated as a function of behavioral report (Bizley et al., 2011; Niwa et al., 2012b). Therefore, our study adds to a growing body of literature that suggests that areas as early as A1 may be contributing to perceptual choices.

Although our findings on choice-related activity in A1 are consistent with those from several other studies, they differ from work previously released from a number of laboratories, including our own, in which choice-related activity is not observed until much later in the ventral pathway. The specific reasons for this discrepancy are unclear, but it seems probable that it may arise from differences between the auditory stimuli and/or the task demands (See Chapter 3 for further discussion).

One caveat to our findings is our selection of neurons. To find neurons, we used a search stimulus of white-noise bursts; and once a putative neuron was isolated, we determined its best frequency. Only neurons that exhibited significant activity to tones during the best-frequency task (relative to a baseline silent period, t-test, $p<0.05$) were tested further. This procedure biased our neural population (1) to be sensitive to white noise and (2) to have best frequencies in our range (400-4000 Hz). Recent studies have reported that different classes of cells in the auditory cortex having different response profiles: for example, pyramidal neurons are more likely to be sharply tuned than interneurons, whereas interneurons are more sensitive to acoustic categories than pyramidal neurons (Moore et al., 2013; Tsunada et al., 2012). It is likely that our neural population included both pyramidal and interneurons; indeed, our sharply tuned neurons were more likely pyramidal cells, whereas the rest of the population may have been a mix of pyramidal and interneurons. Interestingly, our sharply tuned neurons exhibited higher choice probability values than the combined population. This offers an intriguing possibility: perhaps pyramidal neurons are more modulated by choice; or perhaps they play a greater role in the accumulation of

sensory evidence that will contribute to choice or other cognitive processes (Hussar et al., 2012; Mitchell et al., 2007). Further study into the cell-class specific contribution to auditory streaming is merited.

4.3   FUTURE DIRECTIONS

Our use of standard methodologies employed in human audition studies was a crucial addition to the literature, because it showed that monkeys report streaming percepts in a manner consistent with human listeners across a variety of test and control conditions. Because the current studies validated monkeys as a model of human auditory streaming perceptions, future studies can use indirect measures of streaming. This is advantageous, because, as mentioned above, it takes considerable time for the monkeys to learn the streaming task. However, monkeys have been shown to successfully (and relatively quickly) learn several possible alternative tasks. Using a task that monkeys could learn faster and perform with higher accuracy would allow for faster data collection, with potentially a greater number of trials. One such variation of the task would be an oddball detection task in which the monkeys report a stimulus that deviates from the norm (e.g., in intensity or frequency); clever manipulation of the dynamics of the stimulus can be used to test streaming indirectly, by manipulating what is deviant relative to the norm of a given stream. Auditory oddball paradigms have been used successfully with animals in the past (Gifford et al., 2003; Itatani et al., 2014; Mehta et al., 2000; Russ et al., 2008a), and represent a feasible and appealing alternative task to direct reports of streaming.

Further study using the streaming task or alternatives is needed to more completely understand the neural encoding of streaming. First, in order to properly understand how a neuron's tuning affects its activity in this task, it is important to simultaneously record neurons with different best frequencies. This would help elucidate the population response across A1 as

well as identify potential contributions of neurons that are not tuned to tone A (i.e., the best frequency) and the contribution of those neurons whose best frequencies overlap with the tone Bs. Additionally, it would be interesting to compare the neural activity when tone bursts are presented either synchronously or asynchronously. Previous work from Elhilali et al. (2009) only compared this activity in passively listening ferrets, but it is important to examine how active listening affects the temporal coherence of A1 activity.

One other characteristic of sounds that was not manipulated in the current study, but is known to influence the grouping and segregation of acoustic stimuli, is the stimulus' spatial components. Therefore, a version of the task which manipulates the location of the tones (e.g., all tone As came from one location, all tone Bs from another; tones A and B's locations move, or come from random locations) would shine light on contribution of spatial information to early stream formation. Addressing the role of space in stream formation is particularly interesting because spatial information is generally considered characteristic encoded by dorsal auditory pathway. Therefore, examining the effect of space on stream segregation should be done in both dorsal and ventral pathways to further elucidate how these pathways differentially encode information and, potentially, how they communicate (Cohen et al., 2004; Cusack, 2005; Gifford III et al., 2005a; Rauschecker, 2011; Rauschecker, 2012; Rauschecker et al., 2009). For example, when there is a spatial component to stream formation or segregation, is there greater coherence between the activity between the dorsal and ventral auditory pathways?

Finally, the primary auditory cortex is just the beginning of the ventral auditory pathway. Using the same task used in the studies in Chapters 2 and 3, or a variation of the task as described above, the neural responses further along the ventral auditory pathway should be recorded. This will be a key to understanding how these neural signals evolve, and will give further insight into

how auditory stream formation and perception occur.

4.4 REFERENCES

Bizley, J.K., Walker, D.K., Nodal, F.R., King, A.J., Schnupp, J. 2011. Neural correlates of pitch discrimination during passive and active listening, Society for Neuroscience, Vol. Program No. 173.04. 2011 Neuroscience Meeting Planner., Washington, DC.

Bregman, A.S. 1990. Auditory Scene Analysis: The Perceptual Organization of Sound MIT Press, Cambridge, MA.

Christison-Lagay, K.L., Bennur, S., Blackwell, J., Lee, J.H., Schroeder, T., Cohen, Y.E. 2014b. Natural variability in species-specific vocalizations constrains behavior and neural activity. Hear Res 312, 128-42.

Cohen, Y.E., Russ, B.E., Gifford, G.W., 3rd, Kiringoda, R., MacLean, K.A. 2004. Selectivity for the spatial and nonspatial attributes of auditory stimuli in the ventrolateral prefrontal cortex. J Neurosci 24, 11307-16.

Cusack, R. 2005. The Intraparietal Sulcus and Perceptual Organization. J Cogn Neurosci 17, 641-651.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A.J., Shamma, S.A. 2009. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. Neuron 61, 317-29.

Fishman, Y.I., Arezzo, J.C., Steinschneider, M. 2004. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J Acoust Soc Am 116, 1656-70.

Fishman, Y.I., Reser, D.H., Arezzo, J.C., Steinschneider, M. 2001a. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear Res 151, 167-187.

Gifford, G.W., 3rd, Hauser, M.D., Cohen, Y.E. 2003. Discrimination of functionally referential calls by laboratory-housed rhesus macaques: implications for neuroethological studies. Brain, behavior and evolution 61, 213-24.

Gifford III, G.W., Cohen, Y.E. 2005a. Spatial and non-spatial auditory processing in the lateral intraparietal area. Exp Brain Res 162, 509-512.

Hussar, C.R., Pasternak, T. 2012. Memory-guided sensory comparisons in the prefrontal cortex: contribution of putative pyramidal cells and interneurons. J Neurosci 32, 2747-61.

Itatani, N., Klump, G.M. 2014. Neural correlates of auditory streaming in an objective behavioral task. Proc Natl Acad Sci U S A 111, 10738-43.

Izumi, A. 2002. Auditory stream segregation in Japanese monkeys. Cognition 82, B113-B122.

Kuhl, P.K., Miller, J.D. 1975a. SPEECH-PERCEPTION BY CHINCHILLA - VOICED-VOICELESS DISTINCTION IN ALVEOLAR PLOSIVE CONSONANTS. Science 190, 69-72.

Kuhl, P.K., Padden, D.M. 1982. ENHANCED DISCRIMINABILITY AT THE PHONETIC BOUNDARIES FOR THE VOICING FEATURE IN MACAQUES. Perception & Psychophysics 32, 542-550.

Lakatos, P., Musacchia, G., O'Connel, M.N., Falchier, A.Y., Javitt, D.C., Schroeder, C.E. 2013. The Spectrotemporal Filter Mechanism of Auditory Selective Attention. Neuron 77, 750-761.

Ma, L., Micheyl, C., Yin, P., Oxenham, A.J., Shamma, S.A. 2010. Behavioral measures of auditory streaming in ferrets (Mustela putorius). J Comp Psychol 124, 317-30.

Mehta, A.D., Ulbert, I., Schroeder, C.E. 2000. Intermodal selective attention in monkeys. I: distribution and timing of effects across visual areas. Cereb Cortex 10, 343-58.

Micheyl, C., Tian, B., Carlyon, R.P., Rauschecker, J.P. 2005. Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron 48, 139-48.

Mitchell, J.F., Sundberg, K.A., Reynolds, J.H. 2007. Differential attention-dependent response modulation across cell classes in macaque visual area V4. Neuron 55, 131-41.

Moerel, M., De Martino, F., Formisano, E. 2012. Processing of Natural Sounds in Human Auditory Cortex: Tonotopy, Spectral Tuning, and Relation to Voice Sensitivity. J Neurosci 32, 14205-14216.

Moore, A.K., Wehr, M. 2013. Parvalbumin-expressing inhibitory interneurons in auditory cortex are well-tuned for frequency. J Neurosci 33, 13713-23.

Niwa, M., Johnson, J.S., O'Connor, K.N., Sutter, M.L. 2012b. Activity related to perceptual judgment and action in primary auditory cortex. J Neurosci 32, 3193-210.

Noda, T., Kanzaki, R., Takahashi, H. 2013. Stimulus phase locking of cortical oscillation for auditory stream segregation in rats. PLoS One 8, e83544.

Petkov, C.I., O'Connor, K.N., Sutter, M.L. 2003. Illusory sound perception in macaque monkeys. J Neurosci 23, 9155-61.

Petkov, C.I., O'Connor, K.N., Sutter, M.L. 2007. Encoding of illusory continuity in primary auditory cortex. Neuron 54, 153-65.

Rauschecker, J.P. 2011. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hear Res 271, 16-25.

Rauschecker, J.P. 2012. Ventral and dorsal streams in the evolution of speech and language. Frontiers in evolutionary neuroscience 4, 7.

Rauschecker, J.P., Scott, S.K. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12, 718-24.

Recanzone, G.H., Sutter, M.L. 2008. The biological basis of audition. Annu Rev Psychol 59, 119-42.

Russ, B.E., Lee, Y.S., Cohen, Y.E. 2007. Neural and behavioral correlates of auditory categorization. Hear Res 229, 204-12.

Russ, B.E., Orr, L.E., Cohen, Y.E. 2008a. Prefrontal neurons predict choices during an auditory same-different task. Curr Biol 18, 1483-8.

Tsunada, J., Lee, J.H., Cohen, Y.E. 2011a. Representation of speech categories in the primate auditory cortex. J Neurophysiol.

Tsunada, J., Lee, J.H., Cohen, Y.E. 2012. Differential representation of auditory categories between cell classes in primate auditory cortex. J Physiol 590, 3129-39.

5 BIBLIOGRAPHY

Aizenberg, M., Geffen, M.N. 2013. Bidirectional effects of aversive learning on perceptual acuity are mediated by the sensory cortex. Nat Neurosci 16, 994-6.

Andersen, R.A., Bradley, D.C., Shenoy, K.V. 1996. Neural mechanisms for heading and structure-from-motion perception. Cold Spring Harb Symp Quant Biol 61, 15-25.

Anderson, L.A., Christianson, G.B., Linden, J.F. 2009. Stimulus-specific adaptation occurs in the auditory thalamus. J Neurosci 29, 7359-63.

Antunes, F.M., Nelken, I., Covey, E., Malmierca, M.S. 2010. Stimulus-specific adaptation in the auditory thalamus of the anesthetized rat. PLoS One 5, e14071.

Aulanko, R., Hari, R., Lounasmaa, O.V., Naatanen, R., Sams, M. 1993. Phonetic invariance in the human auditory cortex. Neuroreport 4, 1356-8.

Bendor, D., Wang, X. 2005. The neuronal representation of pitch in primate auditory cortex. Nature 436, 1161-5.

Bendor, D., Wang, X. 2006. Cortical representations of pitch in monkeys and humans. Curr Opin Neurobiol 16, 391-9.

Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D. 2004. Neural correlates of sensory and decision processes in auditory object identification. Nature neuroscience 7, 295-301.

Bizley, J.K., Walker, K.M. 2009a. Distributed sensitivity to conspecific vocalizations and implications for the auditory dual stream hypothesis. J Neurosci 29, 3011-3013.

Bizley, J.K., Cohen, Y.E. 2013a. The what, where and how of auditory-object perception. Nat Rev Neurosci 14, 693-707.

Bizley, J.K., Walker, K.M., King, A.J., Schnupp, J.W. 2010. Neural ensemble codes for stimulus periodicity in auditory cortex. J Neurosci 30, 5078-91.

Bizley, J.K., Walker, K.M., Silverman, B.W., King, A.J., Schnupp, J.W. 2009b. Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. J Neurosci 29, 2064-75.

Bizley, J.K., Walker, D.K., Nodal, F.R., King, A.J., Schnupp, J. 2011. Neural correlates of pitch discrimination during passive and active listening, Society for Neuroscience, Vol. Program No. 173.04. 2011 Neuroscience Meeting Planner., Washington, DC.

Bizley, J.K., Walker, K.M., Nodal, F.R., King, A.J., Schnupp, J.W. 2013b. Auditory cortex represents both pitch judgments and the corresponding acoustic cues. Curr Biol 23, 620-5.

Bregman, A.S. 1990. Auditory Scene Analysis: The Perceptual Organization of Sound MIT Press, Cambridge, MA.

Bregman, A.S., Ahad, P.A., Crum, P.A.C., O'Reilly, J. 2000. Effects of time intervals and tone durations on auditory stream segregation. Percept Psychophys 62, 626-636.

Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S., Movshon, J.A. 1996. A relationship between behavioral choice and the visual responses of neurons in macaque MT. Visual Neuroscience 13, 87-100.

Brosch, M., Schreiner, C.E. 1997. Time course of forward masking tuning curves in cat primary auditory cortex. J Neurophysiol 77, 923-43.

Brosch, M., Schulz, A., Scheich, H. 1999. Processing of sound sequences in macaque auditory cortex: response enhancement. J Neurophysiol 82, 1542-59.

Calford, M.B., Semple, M.N. 1995. Monaural inhibition in cat auditory cortex. J Neurophysiol 73, 1876-91.

Carlyon, R.P., Cusack, R., Foxton, J.M., Robertson, I.H. 2001. Effects of Attention and Unilateral Neglect on Auditory Stream Segregation. J Exp Psychol Hum Percept Perform 27, 115-127.

Celebrini, S., Newsome, W.T. 1994. Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. Journal of Neuroscience 14, 4109-24.

Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., Knight, R.T. 2010. Categorical speech representation in human superior temporal gyrus. Nat Neurosci 13, 1428-32.

Christison-Lagay, K.L., Cohen, Y.E. 2014a. Behavioral correlates of auditory streaming in rhesus macaques. Hear Res 309, 17-25.

Christison-Lagay, K.L., Gifford, A.M., Cohen, Y.E. 2015. Neural correlates of auditory scene analysis and perception. International journal of psychophysiology : official journal of the International Organization of Psychophysiology 95, 238-45.

Christison-Lagay, K.L., Bennur, S., Blackwell, J., Lee, J.H., Schroeder, T., Cohen, Y.E. 2014b. Natural variability in species-specific vocalizations constrains behavior and neural activity. Hear Res 312, 128-42.

Coath, M., Brader, J.M., Fusi, S., Denham, S.L. 2005. Multiple views of the response of an ensemble of spectro-temporal features support concurrent classification of utterance, prosody, sex and speaker identity. Network 16, 285-300.

Cohen, M.R., Newsome, W.T. 2009a. Estimates of the Contribution of Single Neurons to Perception Depend on Timescale and Noise Correlation. J Neurosci 29, 6635-6648.

Cohen, M.R., Maunsell, J.H. 2009b. Attention improves performance primarily by reducing interneuronal correlations. Nat Neurosci 12, 1594-600.

Cohen, Y.E. 2012. Auditory Cognition: The Integration of Psychophysics with Neurophysiology. In: Cohen, Y.E., Popper, A.N., Fay, R.R., (Eds.), Neural Correlates of Auditory Cognition, Vol. Springer Handbook of Auditory Research. Springer-Verlag, New York. pp. 1-6.

Cohen, Y.E., Russ, B.E., Gifford, G.W., 3rd, Kiringoda, R., MacLean, K.A. 2004. Selectivity for the spatial and nonspatial attributes of auditory stimuli in the ventrolateral prefrontal cortex. J Neurosci 24, 11307-16.

Cohen, Y.E., Russ, B.E., Davis, S.J., Baker, A.E., Ackelson, A.L., Nitecki, R. 2009c. A functional role for the ventrolateral prefrontal cortex in non-spatial auditory cognition. Proc Natl Acad Sci U S A 106, 20045-50.

Cusack, R. 2005. The Intraparietal Sulcus and Perceptual Organization. J Cogn Neurosci 17, 641-651.

Cusack, R., Deeks, J., Aikman, G., Carlyon, R.P. 2004. Effects of Location, Frequency Region, and Time Course of Selective Attention on Auditory Scene Analysis. Journal of Experimental Psychology: Human Perception and Performance 30, 643-656.

DeWitt, I., Rauschecker, J.P. 2012. Phoneme and word recognition in the auditory ventral stream. Proc Natl Acad Sci U S A 109, E505-14.

Eggermont, J.J. 2001. Between sound and perception: reviewing the search for a neural code. Hear Res, 1-42.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A.J., Shamma, S.A. 2009. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. Neuron 61, 317-29.

Fishman, Y.I., Arezzo, J.C., Steinschneider, M. 2004. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J Acoust Soc Am 116, 1656-70.

Fishman, Y.I., Reser, D.H., Arezzo, J.C., Steinschneider, M. 2000. Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness. J Acoust Soc Am 108, 235-46.

Fishman, Y.I., Reser, D.H., Arezzo, J.C., Steinschneider, M. 2001a. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear Res 151, 167-187.

Fishman, Y.I., Volkov, I.O., Noh, M.D., Garell, P.C., Bakken, H., Arezzo, J.C., Howard, M.A., Steinschneider, M. 2001b. Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. J Neurophysiol 86, 2761-88.

Fritz, J., Shamma, S., Elhilali, M., Klein, D. 2003. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nature neuroscience 6, 1216-23.

Fritz, J.B., David, S.V., Shamma, S.A. 2012. Attention and dynamic, task-related receptive field plasticity in adult auditory cortex. In: Cohen, Y.E., Fay, R.R., Popper, A.N., (Eds.), Neural Correlates of Auditory Cognition. Springer-Verlag. pp. 251-291

Fu, K.M., Shah, A.S., O'Connell, M.N., McGinnis, T., Eckholdt, H., Lakatos, P., Smiley, J., Schroeder, C.E. 2004. Timing and laminar profile of eye-position effects on auditory responses in primate auditory cortex. J Neurophysiol 92, 3522-31.

Gifford, G.W., 3rd, Hauser, M.D., Cohen, Y.E. 2003. Discrimination of functionally referential calls by laboratory-housed rhesus macaques: implications for neuroethological studies. Brain, behavior and evolution 61, 213-24.

Gifford III, G.W., Cohen, Y.E. 2005a. Spatial and non-spatial auditory processing in the lateral intraparietal area. Exp Brain Res 162, 509-512.

Gifford III, G.W., MacLean, K.A., Hauser, M.D., Cohen, Y.E. 2005b. The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. Journal of Cognitive Neuroscience 17, 1471-1482.

Giordano, B.L., McAdams, S., Kriegeskorte, N., Zatorre, R.J., Belin, P. 2012. Abstract encoding of auditory objects in cortical activity patterns. Cereb Cortex.

Gottlieb, Y., Vaadia, E., Abeles, M. 1989. Single unit activity in the auditory cortex of a monkey performing a short term memory task. Exp Brain Res 74, 139-48.

Green, D., Swets, J. 1966. Signal Detection Theory and Psychophysics John Wiley and Sons, Inc, New York.

Griffiths, T.D. 2008. Sensory Systems: Auditory Action Streams? Curr Biol 18, R387-R388.

Griffiths, T.D., Warren, J.D. 2004. What is an auditory object? Nat Rev Neurosci 5, 887-892.

Grimault, N., Bacon, S.P., Micheyl, C. 2002. Auditory stream segregation on the basis of amplitude-modulation rate. J Acoust Soc Am 111, 1340.

Gu, Y., DeAngelis, G.C., Angelaki, D.E. 2007. A functional link between area MSTd and heading perception based on vestibular signals. Nat Neurosci 10, 1038-47.

Gutschalk, A., Micheyl, C., Oxenham, A.J. 2008. Neural Correlates of Auditory Perceptual Awareness under Informational Masking. Plos Biol 6, e138.

Gutschalk, A., Micheyl, C., Melcher, J.R., Rupp, A., Scherg, M., Oxenham, A.J. 2005. Neuromagnetic correlates of streaming in human auditory cortex. J Neurosci 25, 5382-8.

Hackett, T.A. 2011. Information flow in the auditory cortical network. Hear Res 271, 133-46.

Heuer, H.W., Britten, K.H. 2004. Optic flow signals in extrastriate area MST: comparison of perceptual and neuronal sensitivity. J Neurophysiol 91, 1314-26.

Hill, K.T., Bishop, C.W., Yadav, D., Miller, L.M. 2011. Pattern of BOLD signal in auditory cortex relates acoustic response to perceptual streaming. BMC Neurosci 12, 85.

Horvath, J., Czigler, I., Sussman, E., Winkler, I. 2001. Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. Cognitive Brain Research 12, 131-144.

Hussar, C.R., Pasternak, T. 2012. Memory-guided sensory comparisons in the prefrontal cortex: contribution of putative pyramidal cells and interneurons. J Neurosci 32, 2747-61.

Itatani, N., Klump, G.M. 2014. Neural correlates of auditory streaming in an objective behavioral task. Proc Natl Acad Sci U S A 111, 10738-43.

Izumi, A. 2002. Auditory stream segregation in Japanese monkeys. Cognition 82, B113-B122.

Javitt, D.C., Steinschneider, M., Schroeder, C.E., Vaughan, H.G., Jr., Arezzo, J.C. 1994. Detection of stimulus deviance within primate primary auditory cortex: intracortical mechanisms of mismatch negativity (MMN) generation. Brain Res 667, 192-200.

Kaas, J., Hackett, T. 2000. Subdivisions of Auditory Cortex and Processing Streams in Primates. Proceedings of the National Academy of Sciences of the United States of America 97, 11793-11799.

Kaas, J.H., Hackett, T.A. 1999. 'What' and 'where' processing in auditory cortex. Nat Neurosci 2, 1045-1047.

Kajikawa, Y., de La Mothe, L., Blumell, S., Hackett, T.A. 2005. A comparison of neuron response properties in areas A1 and CM of the marmoset monkey auditory cortex: tones and broadband noise. J Neurophysiol 93, 22-34.

Kajikawa, Y., Camalier, C.R., de la Mothe, L.A., D'Angelo, W.R., Sterbing-D'Angelo, S.J., Hackett, T.A. 2011. Auditory cortical tuning to band-pass noise in primate A1 and CM: A comparison to pure tones. Neurosci Res 70, 401-407.

Kikuchi, Y., Horwitz, B., Mishkin, M. 2010. Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. J Neurosci 30, 13021-30.

Kuhl, P.K., Miller, J.D. 1975a. SPEECH-PERCEPTION BY CHINCHILLA - VOICED-VOICELESS DISTINCTION IN ALVEOLAR PLOSIVE CONSONANTS. Science 190, 69-72.

Kuhl, P.K., Miller, J.D. 1975b. SPEECH-PERCEPTION BY CHINCHILLA - PHONETIC BOUNDARIES FOR SYNTHETIC VOT STIMULI. J. Acoust. Soc. Am. 57, S49-S50.

Kuhl, P.K., Padden, D.M. 1982. ENHANCED DISCRIMINABILITY AT THE PHONETIC BOUNDARIES FOR THE VOICING FEATURE IN MACAQUES. Perception & Psychophysics 32, 542-550.

Kusmierek, P., Rauschecker, J.P. 2009. Functional specialization of medial auditory belt cortex in the alert rhesus monkey. J Neurophysiol 102, 1606-22.

Lakatos, P., Musacchia, G., O'Connel, M.N., Falchier, A.Y., Javitt, D.C., Schroeder, C.E. 2013. The Spectrotemporal Filter Mechanism of Auditory Selective Attention. Neuron 77, 750-761.

Law, C.T., Gold, J.I. 2008. Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. Nat Neurosci 11, 505-13.

Leaver, A.M., Rauschecker, J.P. 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J Neurosci 30, 7604-12.

Lee, J.H., Russ, B.E., Orr, L.E., Cohen, Y. 2009. Prefrontal activity predicts monkeys' decisions during an auditory category task. Frontiers in Integrative Neuroscience 3.

Lemus, L., Hernandez, A., Romo, R. 2009. Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. Proc Natl Acad Sci U S A 106, 9471-6.

Logothetis, N.K., Schall, J.D. 1989. Neuronal correlates of subjective visual perception. Science 245, 761-3.

Lu, T., Wang, X. 2000. Temporal discharge patterns evoked by rapid sequences of wide- and narrowband clicks in the primary auditory cortex of cat. J Neurophysiol 84, 236-46.

Ma, L., Micheyl, C., Yin, P., Oxenham, A.J., Shamma, S.A. 2010. Behavioral measures of auditory streaming in ferrets (Mustela putorius). J Comp Psychol 124, 317-30.

McAdams, S., Bregman, A.S. 1979. Hearing musical streams. Compt. Music J. 3, 26-63.

McDermott, J. 2009. The cocktail party problem. Curr Biol 19, R1024-R1027.

Mehta, A.D., Ulbert, I., Schroeder, C.E. 2000. Intermodal selective attention in monkeys. I: distribution and timing of effects across visual areas. Cereb Cortex 10, 343-58.

Merten, K., Nieder, A. 2013. Comparison of abstract decision encoding in the monkey prefrontal cortex, the presupplementary, and cingulate motor areas. J Neurophysiol 110, 19-32.

Mesgarani, N., Chang, E.F. 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485, 233-6.

Micheyl, C., Hunter, C., Oxenham, A.J. 2010. Auditory stream segregation and the perception of across-frequency synchrony. J Exp Psychol Hum Percept Perform 36, 1029-1039.

Micheyl, C., Tian, B., Carlyon, R.P., Rauschecker, J.P. 2005. Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron 48, 139-48.

Micheyl, C., Kreft, H., Shamma, S.A., Oxenham, A.J. 2013a. Temporal coherence versus harmonicity in auditory stream formation. J Acoust Soc Am 133, EL188.

Micheyl, C., Hanson, C., Demany, L., Shamma, S.A., Oxenham, A.J. 2013b. Auditory Stream Segregation for Alternating and Synchronous Tones. J Exp Psychol Hum Percept Perform.

Micheyl, C., Carlyon, R.P., Gutschalk, A., Melcher, J.R., Oxenham, A.J., Rauschecker, J.P., Tian, B., Courtenay Wilson, E. 2007. The role of auditory cortex in the formation of auditory streams. Hear Res 229, 116-31.

Miller, C.T., Dibble, E., Hauser, M.D. 2001. Amodal completion of acoustic signals by a nonhuman primate. Nat Neurosci 4, 783-4.

Mitchell, J.F., Sundberg, K.A., Reynolds, J.H. 2007. Differential attention-dependent response modulation across cell classes in macaque visual area V4. Neuron 55, 131-41.

Moerel, M., De Martino, F., Formisano, E. 2012. Processing of Natural Sounds in Human Auditory Cortex: Tonotopy, Spectral Tuning, and Relation to Voice Sensitivity. J Neurosci 32, 14205-14216.

Moore, A.K., Wehr, M. 2013. Parvalbumin-expressing inhibitory interneurons in auditory cortex are well-tuned for frequency. J Neurosci 33, 13713-23.

Narayan, R., Best, V., Ozmeral, E., McClaine, E., Dent, M., Shinn-Cunningham, B., Sen, K. 2007. Cortical interference effects in the cocktail party problem. Nat Neurosci 10, 1601-7.

Ng, C.W., Plakke, B., Poremba, A. 2013. Neural correlates of auditory recognition memory in the primate dorsal temporal pole. J Neurophysiol.

Nienborg, H., Cumming, B.G. 2009. Decision-related activity in sensory neurons reflects more than a neuron's causal effect. Nature 459, 89-92.

Nienborg, H., Cumming, B.G. 2014. Decision-related activity in sensory neurons may depend on the columnar architecture of cerebral cortex. J Neurosci 34, 3579-85.

Niwa, M., Johnson, J.S., O'Connor, K.N., Sutter, M.L. 2012a. Active engagement improves

primary auditory cortical neurons' ability to discriminate temporal modulation. J Neurosci 32, 9323-34.

Niwa, M., Johnson, J.S., O'Connor, K.N., Sutter, M.L. 2012b. Activity related to perceptual judgment and action in primary auditory cortex. J Neurosci 32, 3193-210.

Niwa, M., Johnson, J.S., O'Connor, K.N., Sutter, M.L. 2013. Differences between Primary Auditory Cortex and Auditory Belt Related to Encoding and Choice for AM Sounds. J Neurosci 33, 8378-95.

Noda, T., Kanzaki, R., Takahashi, H. 2012. Oscillatory neuronal synchronization for auditory stream segregation in auditory cortex of unanesthetized rat. In: Murray, M., (Ed.), 4th Conference on Auditory Cortex, Lausanne, Switzerland. pp. A20.

Noda, T., Kanzaki, R., Takahashi, H. 2013. Stimulus phase locking of cortical oscillation for auditory stream segregation in rats. PLoS One 8, e83544.

O'Connell, M.N., Barczak, A., Schroeder, C.E., Lakatos, P. 2014. Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. J Neurosci 34, 16496-508.

Obleser, J., Eisner, F. 2009. Pre-lexical abstraction of speech in the auditory cortex. Trends Cogn. Sci. 13, 14-9.

Obleser, J., Leaver, A.M., Vanmeter, J., Rauschecker, J.P. 2010. Segregation of vowels and consonants in human auditory cortex: evidence for distributed hierarchical organization. Frontiers in psychology 1, 232.

Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roettinger, M., Eulitz, C., Rauschecker, J.P. 2006. Vowel sound extraction in anterior superior temporal cortex. Hum Brain Mapp 27, 562-71.

Parker, A.J., Newsome, W.T. 1998. Sense and the single neuron: probing the physiology of perception. Annu Rev Neurosci 21, 227-77.

Perrodin, C., Kayser, C., Logothetis, N.K., Petkov, C.I. 2011. Voice cells in the primate temporal lobe. Curr Biol 21, 1408-15.

Petkov, C.I., O'Connor, K.N., Sutter, M.L. 2003. Illusory sound perception in macaque monkeys. J Neurosci 23, 9155-61.

Petkov, C.I., O'Connor, K.N., Sutter, M.L. 2007. Encoding of illusory continuity in primary auditory cortex. Neuron 54, 153-65.

Petkov, C.I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., Logothetis, N.K. 2008. A voice region in the monkey brain. Nat Neurosci 11, 367-74.

Plakke, B., Diltz, M.D., Romanski, L.M. 2013. Coding of vocalizations by single neurons in ventrolateral prefrontal cortex. Hear Res 305, 135-43.

Plakke, B., Hwang, J., Romanski, L.M. 2015. Inactivation of Primate Prefrontal Cortex Impairs Auditory and Audiovisual Working Memory. J Neurosci 35, 9666-75.

Pressnitzer, D., Meddis, R., Delahaye, R., Winter, I.M. 2001. Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus. J Neurosci 15, 6377-6386.

Pressnitzer, D., Sayles, M., Micheyl, C., Winter, I.M. 2008. Perceptual organization of sound begins in the auditory periphery. Curr Biol 18, 1124-8.

Purushothaman, G., Bradley, D.C. 2005. Neural population code for fine perceptual decisions in area MT. Nat Neurosci 8, 99-106.

Quian Quiroga, R., Snyder, L.H., Batista, A.P., Cui, H., Andersen, R.A. 2006. Movement intention is better predicted than attention in the posterior parietal cortex. J Neurosci 26, 3615-20.

Rahne, T., Sussman, E. 2009. Neural representations of auditory input accommodate to the context in a dynamically changing acoustic environment. European Journal of Neuroscience 29, 205-211.

Rauschecker, J.P. 2011. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hear Res 271, 16-25.

Rauschecker, J.P. 2012. Ventral and dorsal streams in the evolution of speech and language. Frontiers in evolutionary neuroscience 4, 7.

Rauschecker, J.P., Tian, B. 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci U S A 97, 11800-6.

Rauschecker, J.P., Tian, B. 2004. Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. J Neurophysiol 91, 2578-89.

Rauschecker, J.P., Scott, S.K. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12, 718-24.

Rauschecker, J.P., Tian, B., Hauser, M. 1995. Processing of complex sounds in the macaque nonprimary auditory cortex. Science 268, 111-4.

Razak, K.A. 2011. Systematic representation of sound locations in the primary auditory cortex. J Neurosci 31, 13848-13859.

Reale, R.A., Brugge, J.F. 2000. Directional sensitivity of neurons in the primary auditory (AI) cortex of the cat to successive sounds ordered in time and space. J Neurophysiol 84, 435-50.

Recanzone, G.H., Sutter, M.L. 2008. The biological basis of audition. Annu Rev Psychol 59, 119-42.

Recanzone, G.H., Schreiner, C.E., Merzenich, M.M. 1993. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. J Neurosci 13, 87-103.

Recanzone, G.H., Guard, D.C., Phan, M.L. 2000. Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. J Neurophysiol 83, 2315-2331.

Romanski, L.M., Averbeck, B.B. 2009. The primate cortical auditory system and neural representation of conspecific vocalizations. Annu Rev Neurosci 32, 315-46.

Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., Rauschecker, J.P. 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci 2, 1131-6.

Russ, B.E., Lee, Y.S., Cohen, Y.E. 2007. Neural and behavioral correlates of auditory categorization. Hear Res 229, 204-12.

Russ, B.E., Orr, L.E., Cohen, Y.E. 2008a. Prefrontal neurons predict choices during an auditory same-different task. Curr Biol 18, 1483-8.

Russ, B.E., Ackelson, A.L., Baker, A.E., Cohen, Y.E. 2008b. Coding of auditory-stimulus identity in the auditory non-spatial processing stream. J Neurophysiol 99, 87-95.

Schebesch, G., Lingner, A., Firzlaff, U., Wiegrebe, L., Grothe, B. 2010. Perception and neural representation of size-variant human vowels in the Mongolian gerbil (Meriones unguiculatus). Hearing research 261, 1-8.

Scheich, H., Brosch, M. 2012. Task-Related Activation of Auditory Cortex. In: Cohen, Y.E., Fay, R.R., Popper, A.N., (Eds.), Neural Correlates of Auditory Cognition. Springer-Verlag. pp. 45-81.

Selezneva, E., Scheich, H., Brosch, M. 2006. Dual Time Scales for Categorical Decision Making in Auditory Cortex. Curr Biol 16, 2428-2433.

Shadlen, M.N., Britten, K.H., Newsome, W.T., Movshon, J.A. 1996. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. Journal of Neuroscience 16, 1486-510.

Shamma, S. 2008. On the Emergence and Awareness of Auditory Objects. PLoS Biol 6, e155.

Shamma, S., Micheyl, C. 2010. Beyond the scenes of auditory perception. Curr Opin Neurobiol 20, 1-6.

Shamma, S.A., Elhilali, M., Micheyl, C. 2011. Temporal coherence and attention in auditory scene analysis. Trends Neurosci 34, 114-23.

Shepard, K.N., Kilgard, M.P., Liu, R.C. 2012. Experience-dependent Plasticity and Auditory Cortex. In: Cohen, Y.E., Fay, R.R., Popper, A.N., (Eds.), Neural Correlates of Auditory Cognition. Springer-Verlag. pp. 293-327.

Shinn-Cunningham, B.G. 2008. Object-based auditory and visual attention. Trends Cogn Sci 12, 182-6.

Singh, P.G. 1987. Perceptual organization of compex-tone sequences: A tradeoff between pith and timbre? J Acoust Soc Am 82, 886-889.

Sinnott, J.M., Petersen, M.R., Hopp, S.L. 1985. Frequency and intensity discrimination in humans and monkeys. J Acoust Soc Am 78, 1977-85.

Steinschneider, M., Arezzo, J.C., Vaughan, J., Herbert G. 1990. Tonotopic features of speech-evoked activity in primate auditory cortex. Brain Res 519, 158-168.

Sussman, E.S. 2005. Integration and segregation in auditory scene analysis. The Journal of the Acoustical Society of America 117, 1285-1298.

Sussman, E.S., Bregman, A.S., Wang, W.J., Khan, F.J. 2005. Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. Cognitive Affective & Behavioral Neuroscience 5, 93-110.

Sussman, E.S., Horvath, J., Winkler, I., Orr, M. 2007. The role of attention in the formation of auditory streams. Perception & Psychophysics 69, 136-152.

Szymanski, F.D., Garcia-Lazaro, J.A., Schnupp, J.W. 2009. Current source density profiles of stimulus-specific adaptation in rat auditory cortex. J Neurophysiol 102, 1483-90.

Taaseh, N., Yaron, A., Nelken, I. 2011. Stimulus-specific adaptation and deviance detection in the rat auditory cortex. PLoS One 6, e23369.

Tian, B., Rauschecker, J.P. 2004. Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. J Neurophysiol 92, 2993-3013.

Tian, B., Reser, D., Durham, A., Kustov, A., Rauschecker, J.P. 2001. Functional specialization in rhesus monkey auditory cortex. Science 292, 290-3.

Tomasello, M. 2008. Origins of Human Communication MIT Press, Cambridge, MA.

Tsunada, J., Lee, J.H., Cohen, Y.E. 2011a. Representation of speech categories in the primate auditory cortex. J Neurophysiol.

Tsunada, J., Lee, J.H., Cohen, Y.E. 2012. Differential representation of auditory categories between cell classes in primate auditory cortex. J Physiol 590, 3129-39.

Tsunada, J., Liu, A.S., Gold, J.I., Cohen, Y.E. 2013. Causal role of primate auditory cortex in auditory perceptual decision-making, Annual Meeting of the Society for Neuroscience, San Diego, CA. pp. 636.01.

Tsunada, J., Liu, A.S., Gold, J.I., Cohen, Y.E. under review. Causal contribution of primate auditory cortex to auditory perceptual decision-making.

Tsunada, J., Baker, A.E., Christison-Lagay, K.L., Davis, S.J., Cohen, Y.E. 2011b. Modulation of cross-frequency coupling by novel and repeated stimuli in the primate ventrolateral prefrontal cortex. Front Psychology 2.

Ulanovsky, N., Las, L., Nelken, I. 2003. Processing of low-probability sounds by cortical neurons. Nat Neurosci 6, 391-398.

Ulanovsky, N., Las, L., Farkas, D., Nelken, I. 2004. Multiple time scales of adaptation in auditory cortex neurons. J Neurosci 24, 10440-10453.

van Noorden, L. 1975. Temporal coherence in the perception of tone sequences. Technical University Eindhoven. pp. 1-127.

Versnel, H., Shamma, S.A. 1998. Spectral-ripple representation of steady-state vowels in primary auditory cortex. The Journal of the Acoustical Society of America 103, 2502-14.

Vliegen, J., Oxenham, A.J. 1999. Sequential stream segregation in the absence of spectral cues. J Acoust Soc Am 105, 339-346.

Wang, X., Lu, T., Bendor, D., Bartlett, E. 2008. Neural coding of temporal information in auditory thalamus and cortex. Neuroscience 154, 294-303.

Wang, X.Q., Merzenich, M.M., Beitel, R., Schreiner, C.E. 1995. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: Temporal and spectral characteristics. Journal of Neurophysiology 74, 2685-2706.

Watkins, P.V., Barbour, D.L. 2011. Level-tuned neurons in primary auditory cortex adapt differently to loud versus soft sounds. Cerebral Cortex 21, 178-190.

Wehr, M., Zador, A.M. 2005. Synaptic mechanisms of forward suppression in rat auditory cortex. Neuron 47, 437-45.

Werner-Reiss, U., Groh, J.M. 2008. A rate code for sound azimuth in monkey auditory cortex: implications for human neuroimaging studies. J Neurosci 28, 3747-3758.

Werner-Reiss, U., Porter, K.K., Underhill, A.M., Groh, J.M. 2006. Long lasting attenuation by prior sounds in auditory cortex of awake primates. Exp Brain Res 168, 272-6.

Winkler, I., Denham, S.L., Nelken, I. 2009. Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci 13, 532-40.

Xu, X., Yu, X., He, J., Nelken, I. 2014. Across-ear stimulus-specific adaptation in the auditory cortex. Frontiers in neural circuits 8, 89.

Yang, S.N., Hwang, H., Ford, J., Heinen, S. 2010. Supplementary eye field activity reflects a decision rule governing smooth pursuit but not the decision. J Neurophysiol 103, 2458-69.

Zar, J. 1996. Biostatistical Analysis Prentice Hall, Upper Saddler River, NJ.

Zhou, Y., Wang, X.Q. 2010. Cortical processing of dynamic sound envelope transitions. J Neurosci 30, 16741-16754.